



```

Db      829  RGVKYPINLQYQKVVYSPHDYQPSVYQPFY-PGFTKESILLQDCWRPNWAYIMEEN 887
QY      275  IAPVNLGFBGFTLSTDTQTLKTLVQLRPTAAGADSFQWTFWNSWNPDSGDTGGILKD 334
Db      888  IAPLLIGWGGHLDGADNEKWKYLRDYII-----ENHIIHTWCFNANSNGDTGGILGVY 941
QY      335  DWQTVDVTVDGYLAP 349
Db      942  DFTWDEKKYSFLKP 956

RESULT 12
QYX3P6
ID      Q9X3P6      PRELIMINARY;      PRT; 1426 AA.
AC      Q9X3P6;
DT      01-NOV-1999 (TrEMBLrel. 12, Created)
DT      01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      CelB.
GN      CelB.
OS      Caldicellulosinruptor sp. Tok7B.1.
OC      Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC      Caldicellulosinruptor.
OC      NCBI_TaxID=80339;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Tok7B.1;
RX      MEDLINE=20171169; PubMed=10706665;
RA      Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA      Bergquist P.B.;
RT      "Multidomain and multifunctional glycosyl hydrolases from the extreme
RT      thermophile Caldicellulosinruptor isolate Tok7B.1.";
RL      Curr. Microbiol. 40:333-340(2000).
CC      -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL
CC      HYDROLASES). AAD30364.1; -.
DR      EMBL; AF078737; AAD30364.1; -.
DR      HSSP; Q06851; INEC.
DR      GO; GO:0003779; F:actin binding; IEA.
DR      GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR      GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR      InterPro; IPR001589; Actbind_actin.
DR      InterPro; IPR001956; CBD_3.
DR      InterPro; IPR008965; Cellul_bind.
DR      InterPro; IPR001000; Glyco_hydro_10.
DR      InterPro; IPR001547; Glyco_hydro_5.
DR      Pfam; PF00942; CBM_3; 3.
DR      PRINTS; PR00134; GLHYDRLASE10.
DR      ProDom; PD001947; CBD_3; 3.
DR      SMART; SM00633; Glyco_10; 1.
DR      PROSITE; PS00019; ACTININ_1; 1.
DR      PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR      PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW      Glycosidase; Hydrolase.
SQ      SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;

Query Match 31.4%; Score 623; DB 2; Length 1426;
Best Local Similarity 35.5%; Pred. No. 1.5e-40;
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY      10  GREITLDANNVPIAGINWFGFTCNVYVHGLWSRDYRSMLDQIKSLGYNITRLPYSDDI 69
Db      1022 GNKLVKDGKPVMTVGTGNTVDFQWCSNLSALATIAKGFMLLPISAEI 1081

QY      70  L---KPGTMFN-SINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRH--RPDCSGQ 123
Db      1082 ILNWSKGYPKPNINY-VYNPELEGLTSLVDFVFKTCVEGLKIMLDIHSKTDAMGH 1140

QY      124  -SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPDP-----ACWGCDDPS 175
Db      1141 IYPWYTDITPFDYKACEWITERYKNDDTIVAFDLKNEPHGKPDQSVFAKWDNSTDI 1200

```

```

QY      176  IDWRLAABEAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWGGNLLQAGQ 222
Db      1201 NNWYAAETCAKRIILAKNPMLIVLEGIEAYPKDDVWTWTSKSSDYISTWGGNLRGVKK 1260
QY      223  YFVVL-NVFNRLVYSAHDYATSVYPTW----FSDPTFPNNMFGIWNKNWGLFNQNIAP 277
Db      1261 YPINLQYQKVVYSPHDYQPSVYQPFY-PGFTKDTLYND---CWRDNWTYIMDNGIAP 1317
QY      278  VWLGEFGITLQSTTDTQTLKTLVQLRPTAAGADSFQWTFWNSWNPDSGDTGGILKDDHQ 337
Db      1318 LLIGEWGGYLDGDNKWKMYLRDYII-----ENHIIHTFCYNANSNGDTGGILGYDFS 1371
QY      338  TVDTVKDGYLAP 349
Db      1372 TWDEQKYNFLKP 1383

RESULT 13
QYX3J2
ID      Q9L3J2      PRELIMINARY;      PRT; 660 AA.
AC      Q9L3J2;
DT      01-OCT-2000 (TrEMBLrel. 15, Created)
DT      01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      1,4-beta-cellobiohydrolase.
DE      CELO.
OS      Clostridium thermocellum.
OC      Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC      Clostridium.
OC      NCBI_TaxID=1515;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=F7;
RA      Zverlov V.V., Schwarz W.H.;
RA      Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
DR      EMBL; AJ275975; CAB76938.1; -.
DR      HSSP; P54583; IECE.
DR      GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR      GO; GO:0016787; F:hydrolase activity; IEA.
DR      GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR      GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR      InterPro; IPR008965; Cellul_bind.
DR      InterPro; IPR002105; Dockerin_1.
DR      InterPro; IPR001547; Glyco_hydro_5.
DR      InterPro; IPR000437; Prok_lipoprot_5.
DR      Pfam; PF00150; cellulase; 1.
DR      Pfam; PF00404; Dockerin_1; 2.
DR      PROSITE; PS00448; CLOS_CELLULOSOME_RPT; 1.
DR      PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR      PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW      Hydrolase.
SQ      SEQUENCE 660 AA; 75247 MW; 04A042002C286CFC CRC64;

Query Match 31.3%; Score 621.5; DB 2; Length 660;
Best Local Similarity 35.8%; Pred. No. 6.9e-41;
Matches 133; Conservative 64; Mismatches 141; Indels 33; Gaps 9;

QY      6  W-HTSGRIILDANNVPIAGINWFGFTCN-VYVHGLWSRDYRSMLDQIKSLGYNITRL 63
Db      201 WLHVVEGLIKDAQGNVTYLTGNTVDFQWCSNLSALATIAKGFMLLPISAEI 260
QY      64  PYSDDI---KPG-TMPSINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRH--RP 118
Db      261 PISAEIILQWNGRVERVTSFVNTYENPRDLGLSLEILDYINENKKNKAMIDMSST 320

QY      119  DCSGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPDP-----DPACWGC 172
Db      321 KDSYQENLWYKNDITMEEFIEAWKIWERYKDDDTVIADVILKNEPHGKYSGNIAKMDDS 380

QY      173  DPSIDWRLAABEAGNAVLNVNPNLLIFVEGVQSYNGDY-----WGGNLLQ 219
Db      381 NDFPNWKKRAAIIABEILAINPNLLIVVEGVEATYPMEGYDYNCGEFTTYCNWGGNLRG 440

```





Db 262 YGPDVFVQSYFNDSNFFNNMFAIWDRFPGFAGSH--ALLLGFEGGKYCEGDARDKVDWQD 313

Qy 298 TLVQYLRLPTAQYAGDSFQWTFWSNPNPSDGTGGILKDDWQT V 339  
||||| : : : ||||| :  
Db 320 ALVKYLR---SKGIN--EGFYWSWNPNSSDGTGGILRDDWTSV 356

RESULT 6

O58925 PRELIMINARY; PRT; 458 AA.

AC O58925; (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-AUG-1998 (TrEMBLrel. 25, Last annotation update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE 458AA long hypothetical endo-1,4-beta-glucanase.

GN PH1171.

OS Pyrococcus horikoshii.

OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae; Thermococcaceae;

CC Pyrococcus.

OX NCBI\_TaxID=53953;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RX MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y., Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y., Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Onfuku Y., Funahashi T., Tanaka T., Kudoh Y., Yamakaki J., Kushida N., Oguchi A., Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K., Masuchi Y., Shizuwa H., Kikuchi H.;

RT "Complete sequence and gene organization of the genome of a hyperthermophilic archaeobacterium, Pyrococcus horikoshii OT3";

RT DNA Res. 5:55-76(1998).

RL EMEL; AP000005; BAA30271.1; -. DR DR PIR; E71059; E71059. DR HSP; P54583; LECE. DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . ; IEA. DR GO; GO:0005975; P:carbohydrate metabolism; IEA. DR InterPro; IPR001547; Glyco\_hydro\_5. DR InterPro; IPR006162; Ppnatne\_S. DR Pfam; PF00150; cellulase; 1. DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_P5; 1. DR PROSITE; PS00012; PHOSPHATANTHEINE; 1. KW Complete proteome. SQ SEQUENCE 458 AA; 51930 MW; CAE48AD43AE8654 CRC64;

Query Match 38.4%; Score 762; DB 17; Length 458;  
Best Local Similarity 45.0%; Pred. No. 2.8e-52;  
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10

Qy 8 TSGRILDANNVPVRIAGINWFQFCNTVVHGLMSRDRSMLDQIKSLGYNTIRLPYS D 67  
||||| : : : ||||| :  
Db 54 TSGE-----THLFGVNWFGETPNHVHGLMKENWEDMLQIKSLGFNAIRLPCT 107  
||||| : : : ||||| :  
Qy 68 DILKPGTGMNSINFRQNWQDLGLTLQWMDXIVAYAGQIGRIITLDHRPDCSQSALW 12  
||||| : : : ||||| :  
Db 108 ESVAPGTPGIDYSK-NPDLRGLDLSQIMEXIIKKAGDLGIFVLVDYHRTGCTIEPLW 16  
||||| : : : ||||| :  
Qy 128 YTSSVSATWISDLQAALQRYKGNPTVVGFDLHNPHD-----PACMGCGDP S 17  
||||| : : : ||||| :  
Db 167 YTEDFSSEDINTWIEVAKFGKNVNVIGADLKNEPHSVTSPPAAVTDTGATWGMNPA 22  
||||| : : : ||||| :  
Qy 176 IDWLAAERAGNAVLSNPMLLIIFVEGVOSYN-----GDSYVWGNIQGAGQYPV 22  
||||| : : : ||||| :  
Db 227 TDWNLAAERIGKAILKVAPHWLIIFVEGTQFTPKTDSSYKNGYNANWGCNLMVAVDYPV- 28  
||||| : : : ||||| :  
Qy 227 LNVF-NRLVTSAHDYATSVPYQWESDPT-FPNNMFGIWNKWGYLFNQIAPYWLGFBG 28  
||||| : : : ||||| :  
Db 286 -NLPRNKLVYSPHYGPDVYNQPFYFGPAKGFPDNLEDIWHFHGVYKLGLGSVVIGBG 34  
||||| : : : ||||| :  
Qy 285 TTL--QSTTDQTKLTLLVOYLRPTAQYAGDSFQWTFWSNPNPSDGTGGILKDDWQT 339  
||||| : : : ||||| :  
Db 345 KYGHGCDPRNVIWONKLVDM--IENKFCDPF---YVSWNPNPSDGTGGILODDWTTI 397  
||||| : : : ||||| :

```

RESULT 7
Q9PF60 PRELIMINARY; PRT; 592 AA.
ID
AC Q9PF60;
AT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DT Endo-1,4-beta-glucanase.
GN XF0818
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
NCBI_TaxID=2371;
RX [1]
RN SEQUENCE FROM N.A.
RP
RC STRAIN=9a5C;
RC MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.B.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.D.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Mitacca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega P.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshakho M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RA "the genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159 (2000).
DR EMBL; AS003921; AAF83628.1; -.
DR PIR; E82759; E82759.
DR HGSP; P54583; LECE.
DR GO; GO:0004573; P:hydrolase activity, hydrolyzing O-glycosyl . . ; IEA.
DR GO; GO:0005955; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac_cellose-bind.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_Hydro.5.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD_II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ
SEQUENCE 592 AA; 59967 MW; 9846DA4EA3B5C89E CRC64;

Query Match 35.0%; Score 695.5; DB 16; Length 592;
Best local similarity 42.6%; Pred. No. 7.7e-47;
Matches 146; Conservative 52; Mismatches 126; Indels 19; Gaps 10
QY 5 YWHTSGREILDANNVPVRIAGINWFGFTTCYVYVHGLWSRDYSMLDQIKSLGVNTRLP 64
DB 26 YSISHGKVDDKGN-QIQLKGGSGFPGFTTNVHGLWTRNKWFITQISGMLNAVRLP 84
QY 65 YSDDIKPGTMTMNSINFRMNOIDLGLTSLQVMDKIYVYAGQICGLRIILDRHRDCSCQS 124
DB 65 FCPASLNSTSPSIDYNR-NPDLOGLSSQLQMDKVKLSDRGIYVLMHHPTDCAIS 143

```

```

Db 165 SELWYTSYPSRMSIDWRMLAKRYASDPTVIGVDLHNEPHCAATWGTGAATTDWRAAAE 224
QY 184 RAGNAVLGVNPNLLIFVEGV-QSYNGDSYWMGNLQAGQVPPVNLVNPRLVYSAHDYAT 242
Db 225 RGNNAVLAENPKLLVLEGIDHQADGTCTWGGALDSATASVRLTVANRVVYSPHDYPS 284
QY 243 SVYPTWFSDDTPFPNNPFGINWKNWGYLFNQIAPVWLGEFTTLQSTDTQTLTKTLVQY 302
Db 285 TIYGQWFSASNYPTNLPGIWDHAGYLAQKDIAPVLVGEFTTLETASDKQWMLTLVGY 344
QY 303 LRPTAQYAGDSFQWTFWSNPDSDGTGGLAKDDQWTDVTKDGYLAPI 350
Db 345 LSST---GISS---SFMFNPDSGDTGGIVAKSDWVTPEAKLDLALAPI 386

RESULT 4
Q9EVR6 PRELIMINARY; PRT; 518 AA.
AC Q9EVR6
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Extracellular endoglucanase (ENGXCA protein) precursor.
GN ENGXCA
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRLL-B1459;
RA Schroeter K., Puhler A., Becker A.;
RT "engxca major extracellular endoglucanase.";
RL Submitted (DEC-2000) to the ENBL/GenBank/DBJ databases.
DR EMBL; AJ304415; CAC18529.1; -.
DR HSSP; P54583; 1ECC.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008965; Cellul bind.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW SIGNAL.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 518 EXTRACELLULAR ENDOGLUCANASE (ENGXCA
PROTEIN).
SQ SEQUENCE 518 AA; 55588 MW; 9A0FF67BEB712C CRC64;

Query Match 40.5%; Score 803.5; DB 2; Length 518;
Best Local Similarity 46.9%; Pred. No. 1.6e-55;
Matches 161; Conservative 61; Mismatches 102; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRLP 64
Db 26 YSINNSRQIVDDSGKVQLGKGVNVFGFETGNVHVGHWARNKEMINQMGGLGFNAVRLP 85
QY 65 YSDILKPGTWPNINFRQMNQDLQGLTSLQVMDKIVAVAGQIGLRIILDRHPPDCSGQS 124
Db 86 FCPATLRSDTNPASIDYSR-NADLQGLTSLQILDKVAEFGNAGWVYLLDHTHPDCAGIS 144
QY 125 ALWYTSVSEATWISDLOALAQRYKGNPTVVGFPLHNEPHDPCWCGDPSIDWRLAER 184
Db 145 ELWYTSYSEAQWLADLRFRVANRYKNVYVGLGLDKNEPFGAATWGTGNAATDWNKAAER 204
QY 185 AGNAVLSPNPLLIIFVEGVQ-----SYNGDSYWMGNLQAGQVPPVNLVNP-NRLVYSAH 238
Db 205 GSAVLAVAPKWLAVSGITDNPVCSITNG-GIFWGNLQPLACTP--LNPANRLLLAHP 261
QY 239 DYATSVPTQWFSDDTPFPNNPFGINWKNWGYLFNQIAPVWLGEFTTL--QSTTDQTLW 296
Db 262 VYGPDVVFQSYFNSDNPNNPAAIWERHFGQFAGTH--ALLLGEFGKGEGDARDKTWQ 319
QY 297 KTLVQYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDQWTV 339

```

```

Db 320 DALVKYLR---SKGIN--QGFYWSNPNNSGDTGGILRDDWTSV 357

RESULT 5
Q8PPS3 PRELIMINARY; PRT; 474 AA.
AC Q8PPS3
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Cellulase.
GN ENGXCA OR XAC0612.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camarvan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Senna J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.C., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB011689; AM35501.1; -.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac cellose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001547; Glyco_Hydro_5.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 51294 MW; 379B84D63CASA31B CRC64;

Query Match 40.4%; Score 802.5; DB 16; Length 474;
Best Local Similarity 46.2%; Pred. No. 1.7e-55;
Matches 158; Conservative 63; Mismatches 103; Indels 19; Gaps 9;

QY 6 WHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRLPY 65
Db 26 YSVSNKIKVDNNGVQLKGVNVFGFETGNVHVGHWARNKEMINQMGGLGFNAVRLP 85
QY 66 SDDILKPGTWPNINFRQMNQDLQGLTSLQVMDKIVAVAGQIGLRIILDRHPPDCSGOSA 125
Db 86 CPATLRSDTNPASIDYSR-NADLQGLTSLQILDKVAEFGNAGWVYLLDHTHPDCGALISE 144
QY 126 LWYTSVSEATWISDLOALAQRYKGNPTVVGFPLHNEPHDPCWCGDPSIDWRLAERA 185
Db 145 LWYTSYSEAQWLADLRFRVANRYKNVPSVIGVDLKNPFGAATWGTGNAATDWNATAERG 204
QY 186 GNAVLSPNPLLIIFVEGVQ-----SYNGDSYWMGNLQAGQVPPVNLVNP-NRLVYSAH 239
Db 205 SAAVLAVAPKWLAVSGITDNPVCSITNG-GIFWGNLQPLACTP--LNPANRLLLAHPV 261
QY 240 YATSVPTQWFSDDTPFPNNPFGINWKNWGYLFNQIAPVWLGEFTTL--QSTTDQTLWK 297

```

```

QY 124 SALWTTSSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 156 SELWTTSSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 215
QY 184 RAGNAVLSVNPGLLIFVEGVQ---SYNGDSYWMGNLQAGQYPPVVLNPNRLVYSAHDY 240
Db 216 RAGNAVLSVNPGLLIFVEGVQ---SYNGDSYWMGNLQAGQYPPVVLNPNRLVYSAHDY 275
QY 241 ATSVYPTQWFSDFTPNNMFGIWNKNGYLFNQNTAPVWLGBFG--TTLQSTTDQTLWLT 298
Db 276 GPGVSSQPFNDSTFSPNLPAIWDQWYISKNTAPVLVGFGRNVDLSPEGKQWNA 335
QY 299 LVQYLPRPTAQYAGDSQWTFMSWNPDSGDTGILKDDWQTVTDVKGYLAPIKSIFFDPV 358
Db 336 LVQYLPRPTAQYAGDSQWTFMSWNPDSGDTGILKDDWQTVTDVKGYLAPIKSIFFDPV 385

RESULT 2
Q9K5C7 PRELIMINARY; PRT; 745 AA.
AC Q9K5C7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE Cellulase precursor (Fragment).
GN CELA.
OS Clavibacter michiganensis.
OC Plasmid pCML.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=28447;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Meletzus D.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-614 FROM N.A.
RC STRAIN=NCPPB 382;
RX MEDLINE=20331586; PubMed=10875331;
RA Jahr H., Dreier J., Meletzus D., Bahro R., Eichenlaub R.J.;
RT "The endo-beta-1,4-glucanase CelA of Clavibacter michiganensis subsp.
RT michiganensis is a pathogenic determinant required for induction of
RT bacterial wilt of tomato.";
RL Mol. Plant Microbe Interact. 13:703-714 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Jahr H.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; X62582; CAA44467.2; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; BacCelose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW SIGNAL.
FT CHAIN 1 66 POTENTIAL.
FT NON_TER 67 >745 CELLULASE.
FT SEQUENCE 745 AA; 77886 MW; 584501.5EB8D4F27C CRC64;

Query Match 45.8%; Score 909.5; DB 2; Length 745;
Best Local Similarity 50.3%; Pred. No. 1e-63;
Matches 175; Conservative 51; Mismatches 111; Indels 11; Gaps 4;

```

```

QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRL 63
Db 70 GMLHTAGKIVTASGAPTYTIGLAWFQVETSSCAPHGLDTITLAGGQVHILKQGFVTRL 129
QY 64 PYSDDLKPGTVPNSINFRQMDLOGLTSLQVMDKIVAYAGQIGRLIILDRHRPDCSGQ 123
Db 130 PFSNOCL----AASGVTVGSADPSLAGLTPLQVMDHVVASAKSAGLDVILDOHRPDSGGQ 185
QY 124 SALWTTSSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 186 SELWTTSSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 245
QY 184 RAGNAVLSVNPGLLIFVEGVQ---SYNGDSYWMGNLQAGQYPPVVLNPNRLVYSAHDYAT 242
Db 246 RGNNAVLAENPNLLIVVEGIDHEDGSGTWMGALGLVGNAPVRLSVANRVVYSPHDYPS 305
QY 243 SVYPTQWFSDFTPNNMFGIWNKNGYLFNQNTAPVWLGBFGTTLQSTTDQTLWLTQVY 302
Db 306 TIYQSWFSASNYPANLPGIWDHAGWYLAKKDAPVLVGFGRNVDLSPEGKQWNA 365
QY 303 LRPTAQYAGDSQWTFMSWNPDSGDTGILKDDWQTVTDVKGYLAPI 350
Db 366 LSST---GISS---SFWAFNPNSGDTGGIVKSDWVTPEQAKLDALAPI 407

RESULT 3
Q9AF65 PRELIMINARY; PRT; 727 AA.
AC Q9AF65;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE Cellulase CelA.
OS Clavibacter michiganensis (subsp. sepedonicus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=31964;
RN [1]
RP SEQUENCE FROM N.A.
RA Laine M., Haapalainen M., Wahloors T., Kankare K., Nissinen R.,
RA Kassuwi S., Metzler M.C.;
RT "The cellulase encoded by the native plasmid of Clavibacter
RT michiganensis subsp. sepedonicus plays a role in virulence and
RT contains an expansin-like domain.";
RL Physiol. Mol. Plant Pathol. 0:0-0 (2001).
DR EMBL; AY007311; AAK16222.1; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR SEQUENCE 727 AA; 75811 MW; 96DF9B664873985D CRC64;

Query Match 45.4%; Score 902.5; DB 2; Length 727;
Best Local Similarity 50.3%; Pred. No. 3.5e-63;
Matches 175; Conservative 48; Mismatches 114; Indels 11; Gaps 4;

QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRL 63
Db 49 GMLHTAGKIVTASGAPTYTIGLAWFQVETSSCAPHGLDTITLAGGQVHILKQGFVTRL 108
QY 64 PYSDDLKPGTVPNSINFRQMDLOGLTSLQVMDKIVAYAGQIGRLIILDRHRPDCSGQ 123
Db 109 PFSNOCL----AASGVTVGSADPSLAGLTPLQVMDHVVASAKSAGLDVILDOHRPDSGGQ 164
QY 124 SALWTTSSVSEATWISDLOALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:29 ; Search time 36.3333 Seconds  
(without alignments)  
3108.867 Million cell updates/sec

Title: US-09-997-504A-14  
Perfect score: 1986  
Sequence: 1 AGGGVHTSGREILDANNVP.....VDTVKDGYLAPIKSIIDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_todent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_ivirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1049	52.8	397	2 Q8RP23	Q8RP23 paenibacill
2	909.5	45.8	745	2 Q9K5C7	Q9K5C7 clavibacter
3	902.5	45.4	727	2 Q9AF65	Q9AF65 clavibacter
4	803.5	40.5	518	2 Q9EVR6	Q9EVR6 xanthomonas
5	802.5	40.4	474	16 Q8PPS3	Q8PPS3 xanthomonas
6	762	38.4	458	17 Q58925	Q58925 pyrococcus
7	695.5	35.0	592	16 Q9P60	Q9P60 xylella fas
8	688.5	34.7	514	17 Q9V052	Q9V052 pyrococcus
9	681	34.3	482	16 Q97KK6	Q97KK6 clostridium
10	656.5	33.1	614	16 Q87AH4	Q87AH4 xylella fas
11	629.5	31.7	1000	2 Q24820	Q24820 thermophili
12	623	31.4	1426	2 Q9X3P6	Q9X3P6 caldicellul
13	621.5	31.3	660	2 Q9L3J2	Q9L3J2 clostridium
14	602.5	30.3	534	2 Q9EYQ0	Q9EYQ0 clostridium
15	583	29.4	616	2 Q7X2N2	Q7X2N2 thermomonos
16	524.5	26.4	425	2 Q93Q07	Q93Q07 thermus cal

17	492	24.8	574	2 Q9S3V3	Q9S3V3 cellulomona
18	490	24.7	341	2 P96310	P96310 anaerocellu
19	393	19.8	630	2 Q8RJY7	Q8RJY7 stigmatella
20	324	16.3	565	16 Q9PF68	Q9PF68 xylella fas
21	311.5	15.7	569	16 Q87AG9	Q87AG9 xylella fas
22	272	13.7	508	10 Q9LTM8	Q9LTM8 arabidopsis
23	271.5	13.7	590	16 Q8P513	Q8P513 xanthomonas
24	257.5	13.0	522	10 Q9SAB6	Q9SAB6 arabidopsis
25	210	10.6	526	10 Q9LFS2	Q9LFS2 arabidopsis
26	206.5	10.4	488	10 Q9LFE7	Q9LFE7 arabidopsis
27	194.5	9.8	354	16 Q8PEF2	Q8PEF2 xanthomonas
28	190.5	9.6	555	10 Q7XUQ4	Q7XUQ4 oryza sativ
29	188.5	9.5	561	10 Q9LTN0	Q9LTN0 arabidopsis
30	184.5	9.3	437	3 Q99036	Q99036 trichoderma
31	176.5	8.9	377	3 Q90012	Q90012 aspergillus
32	170.5	8.6	694	2 Q9Z187	Q9Z187 bacillus st
33	169	8.5	439	3 Q92401	Q92401 agarius bi
34	168	8.5	439	3 Q9F893	Q9F893 agarius bi
35	166.5	8.4	357	16 Q8PRD3	Q8PRD3 xanthomonas
36	166	8.4	377	16 Q8PRD3	Q8PRD3 xanthomonas
37	164.5	8.3	558	10 Q7XUQ6	Q7XUQ6 oryza sativ
38	163	8.2	364	2 Q9F0G8	Q9F0G8 rhizobium m
39	160.5	8.1	516	2 Q66185	Q66185 bacillus ci
40	160	8.1	550	10 Q7XUQ5	Q7XUQ5 oryza sativ
41	154.5	7.8	722	16 Q97L56	Q97L56 clostridium
42	153.5	7.7	363	2 Q07652	Q07652 cellvibrio
43	153.5	7.7	1449	10 Q8RU51	Q8RU51 oryza sativ
44	153	7.7	357	2 Q59441	Q59441 fibrobacter
45	151.5	7.6	356	16 Q9PA12	Q9PA12 xylella fas

## ALIGNMENTS

## RESULT 1

ID	Q8RP23	PRELIMINARY;	PRT;	397 AA.
AC	Q8RP23;			
DT	01-JUN-2002 (TREMREL. 21, Created)			
DT	01-JUN-2002 (TREMREL. 21, Last sequence update)			
DT	01-JUN-2003 (TREMREL. 24, Last annotation update)			
DE	Endo-beta-1,4-glucanase			
OS	Paenibacillus sp. KCTC848P.			
OC	Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.			
OX	NCBI_TaxID=109199;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=KTC 8848P;			
RA	Park J.N., Kim H.O., Shin D.J., Lee H.B., Chun S.B., Bai S.;			
RT	"Cloning of Paenibacillus sp. endo-beta-1,4-glucanase gene and its co-			
RT	expression with Endonocyes fibuliger beta-glucosidase gene in			
RT	Saccharomyces cerevisiae."			
RL	Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL; AF345984; AAL8749.1; ..			
DR	GO: GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.			
DR	GO: GO:0005975; P:carbohydrate metabolism; IEA.			
DR	InterPro; IPR001547; Glyco_hydro_5.			
DR	Fram; PF00150; cellulase; 1.			
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.			
SQ	SEQUENCE 397 AA; 44539 MW; 1736A4C0C04E298F CRC64;			
Query Match 52.8%; Score 1049; DB 2; Length 397;				
Best Local Similarity 54.7%; Pred. No. 3.3e-75;				
Matches 197; Conservative 44; Mismatches 103; Indels 16; Gaps 5;				
QY	4 GYVHTSGREILDANNVPVRIAGINWFGFTCNVYVHGLWSRDYRMLDQIKSLGNVTRL 63			
DB	37 GYVHTCGNKIVDETCKEAFNGLNWFGLETPNYTLHGLMSRSMDDMLDQVKKEGYNLRL 96			
QY	64 PYSDDLKPGTNPINFRMNQDLQGLTSLQVMDKIVAYAGIQLRLILDRHRPDCSGQ 123			
DB	97 PISNQLFDSSSRADSIDYK-NPDVLVGLTPTQIMDKLIEKAGQGIQLDRHRPDCSGQ 155			



*This Page Blank (uspto)*

Db 402 QVWDSLSRGFTVITINSHDD-----WIKEDYNGNIERFEKIW-----EQIAERFKNKS 450  
QY 153 TVVGFDLHNEHDPACWCGDPSIDWLAARAGNAVLSVNPVLLIFVEGVQSYNGDSYV 212  
Db 451 ENLLFEIMNEPFG-----NITDEQID--DMNSRIILKIIRKTNPRIVII-----GGYV 497  
QY 213 WGNLQAGAGQPVVVLN--VPN--RLVSAHDYATSVYPQWFSDD--TFPNNMFGIWN--- 264  
Db 498 -----NSYNTLVNIKIPDPYLICTPHY-----DPYBFTHKWGTWGTQE 538  
QY 265 -----KNGVLYFNQNIAPVWLGRTTQSTTQTLKTLVQLRPTAQVAD 312  
Db 539 DMTVTVRVDFVKWS-----DKNIPVYGF--AVMAYADRTSRVWYDFISDAAL--ER 591  
QY 313 SFQWTFMSWNPDSGDTGILKD-----DMQVDTVRKDGVLAPIKSSIFDP 357  
Db 592 GFACSVW-----DNGVFGSLDNDMAIYNRDTTFTD-----ETLNALFNP 631

RESULT 15  
GUNE\_RUMAL STANDARD; PRT; 409 AA.  
AC 233661;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-NOV-1991 (Rel. 20, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase) (EGB).  
GN CELB.  
OS Ruminococcus albus.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;  
OC Ruminococcus.  
OX NCBI\_TaxID=1264;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SY3;  
RX MEDLINE=91066833; PubMed=22506649;  
RA Poole D.M., Hazlewood G.P., Laurie J.I., Barker P.J., Gilbert H.J.;  
RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase  
RT genes celA and celB.";  
RL Mol. Gen. Genet. 223:217-223(1990).  
CC -! CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -! SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
CC hydrolases).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X54932; CAA38693.1; -.  
CC PIR; S12018; S12018.  
CC HSP; P17901; 1EDG.  
DR InterPro; IPR001547; Glyco\_hydro.5.  
DR InterPro; IPR000437; Prok\_lipprot\_s.  
DR Pfam; PF00150; cellulase; 1.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; UNKNOWN\_1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 409 ENDOGLUCANASE B.  
FT ACT\_SITE 212 212 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 332 332 NUCLEOPHILE (BY SIMILARITY).  
SQ SEQUENCE 409 AA; 45523 MW; 119081DFA3BFDD54 CRC64;

Query Match 6.7%; Score 134; DB 1; Length 409;  
Best Local Similarity 24.1%; Pred. No. 0.0027;  
Matches 78; Conservative 44; Mismatches 119; Indels 82; Gaps 16;

QY 49 MLDQIKSLGYNTIRLPYS-----DDILKPGTNPNSINFRQNMQDLQGLTSLQVMDKI 100  
Db 112 MIDMLPEAGFNVLRIPVSWGNHLLIDNNYTIIDPAWMDR-----VQEI 152  
QY 101 VAYAGQIGLRILDRHRPDCSQSALWYTSVSEATW-ISDLQA-----LAQRYKGNPTVW 155  
Db 153 VNYGIDDDGMVYLINTHHEE-----WMPKPEKXGDIEELKAIWSQIADRFKGYDEHL 205  
QY 156 GFDLHNEHDPACWCGDPSIDWRLAERA-----GNAVLSVNPVNPVLLIFVEGVQV--YN 207  
Db 206 IFEGLNEP-----RLRGEAEWTGTSEAREIINEYEKAFVETVRASGNN 250  
QY 208 GDSYVW-----CGNLQAGQVPLVNVNPNLVYSAHDYAT-SVYPQWFSPTPEPN 258  
Db 251 GDRCLMITGYAASSGYNLSAIELP-----EDSDKLLIISVHAYLPYSFALDTKGTDKYDPE 307  
QY 259 --MPGIMWKNWGYLFNQNIAPVWLGEFGTTLQSTTDTQTLKTLVQLRPTAQYGADSFQW 316  
Db 308 TAIPTLF-ESLNELFISRDIPVIVGEFGSMKNKNIDDR-VKCLDDYLGNAAKYDIPCVW 365  
QY 317 TFSWNPDSGDTGGIL-----KDDW 336  
Db 366 DNYA-RIGNGENFGLLRQYDW 387

Search completed: August 2, 2004, 16:34:57  
Job time : 9.33333 secs



[illegible]

```

3-1.2 F-EHNLNFIKSGFVTVGFGTMMKNENTEDR-VKLEDEYLAATAAKYDIPCVWMDNYA-R 368
QY 323 PDSGDTGGILKD---DWQTVDITVK 343
DB 369 IGCNGENFGLMNRADLEWYFFDLIE 392

RESULT 13
GUND_CLOCE STANDARD; PRT; 584 AA.
ID GUND_CLOCE
AC P25472;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase D)
DE (Cellulase D) (EGCCD).
DE CELCCD.
OS Clostridium cellulolyticum.
OS Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
XX [1]_TaxID=1521;
RN [1]_SEQUENCE FROM N.A.
RP SEQUENCE ATCC 35319;
RX MEDLINE=92009193; PubMed=1916275;
RA Shima S., Igarashi Y., Kodama T.;
RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,
RT celCCD, of Clostridium cellulolyticum.";
RL Gene 104:33-38(1991).
CC -! FUNCTION: The biological conversion of cellulose to glucose
CC Generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glycosidic bonds; (2)
CC Exocellulohydrolases that cut the disaccharide cellobiose
CC from the nonreducing end of the cellulose polymer chain; (3)
CC Beta-1,4-glicosidases which hydrolyze the cellobiose and other
CC short cello-oligosaccharides to glucose.
CC -! CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -! PATHWAY: Cellulose degradation.
CC -! DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS

```

[illegible]

Db	274	ASNVPFG-NAWTG-AHSWNQWYTFNGVTVMKGINDPGHQLVFEVHQYLDGDSGQSA	331
Qy	241	--ATSVYPTQWFGSDPTFPNNKPGIMKNWYLFNQNIAPWVLGEFGTTLQSTTDQWLKT	298
Db	332	NCVSATIGAQRLOQFT-----TWLRSNGY-----RGFLGEFGAASNDTCNQA-VSN	376
Qy	299	LVQYLPTAQYGADSGFQWTFWSNPDSDGTGGILK--DDMQTVDYDKDGYLAP	349
Db	377	MLTFVKNNADV-----WTCWAWAGPWGGYMYSIENPGVYDKFQMSVLAP	423
RESULT 11			
GUNE_CLOTH			
ID	GUNE_CLOTH	STANDARD;	PRT; 814 AA.
AC	P10477;		
DT	01-JUL-1989	(Rel. 11, Created)	
DT	01-JUL-1989	(Rel. 11, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	Endoglucanase E precursor (EC 3.2.1.4) (EGE)	(Endo-1,4-beta-glucanase)	
DE	(Cellulase E).		
GN	CLE.		
OS	Clostridium thermocellum.		
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;		
OC	Clostridium.		
OC	NCBI_TaxID=1515;		
[1]			
RN	SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.		
RX	MEDLINE=69137992; PubMed=3066698;		
RA	Hall J., Hawlewood G.P., Barker P.J., Gilbert H.J.;		
RT	"Conserved reiterated domains in Clostridium thermocellum		
RT	endoglucanases are not essential for catalytic activity."		
RL	Gene 69:29-38(1988).		
CC	!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-		
CC	GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-		
CC	GLUCANS.		
CC	!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic		
CC	linkages in cellulose, lichenin and cereal beta-D-glucans.		
CC	!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS		
CC	WELL AS IN OTHER C.THERMOCCELLUM CELLULOSONOME ENZYMES. THIS DOMAIN		
CC	MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.		
CC	!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl		
CC	hydrolases).		
-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>		
CC	or send an email to <a href="mailto:license@sb-sib.ch">license@sb-sib.ch</a> ).		
-----			
DR	EMBL; M22759; AAA23224.1; --		
DR	PIR; J03447; CZCLEM.		
DR	HSPF; P17901; LEDG.		
DR	InterPro; IPR002105; Dockerin 1.		
DR	InterPro; IPR001547; Glyco_hydro 5.		
DR	InterPro; IPR001087; Lipase_GDSL.		
DR	Pfam; PF00150; Cellulase; 1_		
DR	Pfam; PF00404; Dockerin; 1_		
DR	Pfam; PF00685; Lipase_GDSL; 1.		
DR	PROSITE; PS00448; CLOS_CELLULOSONOME_RPT; 2.		
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.		
KW	Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.		
FT	SIGNAL	1 34	
FT	CHAIN	35 814	ENDOGLUCANASE E.
FT	ACT_SITE	193 193	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	316 316	NUCLEOPHILE (BY SIMILARITY).
FT	DOMAIN	415 474	2 X 24 AA APPROXIMATE REPEATS.
FT	REPEAT	415 438	1.
FT	REPEAT	451 474	2.
SO	SEQUENCE	814 AA; 90244 MW; C6FA24B8D1523632 CRC64;	

FT DOMAIN 281 748 CATALYTIC.  
 FT DISULFID 39 133 BY SIMILARITY.  
 FT ACT\_SITE 503 503 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 653 653 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 748 AA; 80197 MW; AC96104137932B76 CRC64;  
 Query Match 15.0%; Score 298.5; DB 1; Length 748;  
 Best Local Similarity 23.4%; Pred. No. 5.5e-16;  
 Matches 111; Conservative 61; Mismatches 138; Indels 165; Gaps 23;  
 QY 7 HTSREILDANNVVRAGINWFGFTCN-----YVHGLW-----S 43  
 Db 290 NTKGLTKDGLLPARGC--NWFLGRGHEFSNDADNPSGAPMELVAGNMWVNNQSGS 347  
 QY 44 RYRSMDLQIKSLGYNTIRLPYSDDL---KPGTMPSINFRQMDLQGLTSLQVMDK 99  
 Db 348 RTIOQTWELKQOITWRLPIAQTLDANDPQSRNLKXHQSIQS-----NARQALD 403  
 QY 100 IVAYAGQIGRIIDRH-----RPDCSGQALWY-----TSSVS 133  
 Db 404 FIKLADQNDIQIFDIHSCSNVYVGRAGRLDARPPYVDANRVGYDFTREYSCSATNPS 463  
 QY 134 EAT-----WLSDLQALA--QRYKGNPVVGVFDLHNEPHDPACWCGDPSIDWRLAA 182  
 Db 464 SVTRIHYDKQKWLANUREIAGLSAKLVNIGIDVFNFPYD-YTWA-----EHWGMV 516  
 QY 183 ERAGNAVLSVNPMLLIIFVEGVQSYNGDSY-----WKGNGLQAGQY 223  
 Db 517 BEAQAINNEVPNMLIIVEGI-SANANTQDGTPTDTSVPEPHGSTDLNPNWGNLYEAGAN 575  
 QY 224 PVVLNVP-NRLVYSADHYATSVYQTFWSDPT-----FPNNMPG 261  
 Db 576 PP--NIPKRLLFSPHTYGPVSFVQRFMDPAQTECAGLEGDEAAQAARCRIVNPTVLEQ 633  
 QY 262 IWNKNGVLFNQNIAPVWLGFBG-----TTIQSTTDQTLTKLTVQYL 303  
 Db 634 GWSEHFGYLRGLVG--ILIGFGGNWDPGAKSSQADENAWSHITTVNDQWQQAASYF 692  
 QY 304 RPTAQYAGDSQWTFWNPDSGTGILKDDWQTVTDVKGYLAKPSIFBPV 358  
 Db 693 K---RKGINA---CYWSNPNESADTMG-----YLTTP-----WDPV 722

## RESULT 9

GUN1\_RALSO STANDARD; PRT; 424 AA.  
 AC P58599;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN EGL OR RSP0162 OR RS05516.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RC MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artigianave F., Guzy J., Mangerot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunne S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguer P., Thebault P., Whalen P., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RA "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

CC (Probable).  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AL646076; CAD17313.1; ALT\_INIT.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC InterPro; IPR000437; Prok\_hydroprot\_5.  
 CC Pfam; PF00150; cellulase; 1.  
 CC PROSITE; P500013; PROKAR LIPOPROTEIN; 1.  
 CC PROSITE; P500659; GLYCOSYL HYDROL F5; 1.  
 CC Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;  
 CC Lipoprotein; Signal; Plasmid; Complete proteome; Palmitate.  
 KW SIGNAL 1 19  
 FT PROPEP 20 43 BY SIMILARITY.  
 FT CHAIN 44 424 S-diacylglycerol cysteine (By  
 FT LIPID 20 20 similarity).  
 FT LIPID 20 20 N-palmitoyl cysteine (By similarity).  
 FT ACT\_SITE 247 247 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 359 359 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 424 AA; 45080 MW; 5BF0D161AB7C87EF CRC64;  
 Query Match 9.3%; Score 184; DB 1; Length 424;  
 Best Local Similarity 25.5%; Pred. No. 3.1e-07;  
 Matches 86; Conservative 50; Mismatches 125; Indels 76; Gaps 19;  
 QY 35 NTVVHGLWSRDYRSLMDQIKSLGYNTIRLPYSDDLKKEGTPMNSINFRQMDQ--DLQGLT 92  
 Db 139 NYIYFSADSATY-----YKNGMNLVRLPFRWERLQ-----TLNQALDANELS 182  
 QY 93 SLQVMDKIVAYAGQIGLRIIDRHDPDCSGQALWY-----TSSVSEATWISDIQALQR 147  
 Db 183 RUTGFVNAVTAAGQ--TVLLDPH-----NYARYYGNVIGSSAVPSAYADFWRVATQ 233  
 QY 148 YKGNPTVYGVFDLHNEPHDPACWCGDPSIDWRLAAERAGNAVLSVNPMLLIIFVEGVQSYN 207  
 Db 234 FKGNARVI-FGLMNEPNSM-----PTEQMLSGANAALAAIRSANASNVVFPVG-NAWT 284  
 QY 208 GDSYVWGGNLQAGQYPVV--LVVPSNR-LVYSAHDY-----ATSVYPTWPSD--PT 254  
 Db 285 G-AHSNQNQWYGTNGTVKNGINDPGRNLVFEVHQYLDGDSGQSGASCSVSATIGERLQD 343  
 QY 255 FPNMMPGIMWKNWGYLFNQNIAPVWLGFBGTTTQSTTDQTLTKLTVQYLRTAQYAGDSF 314  
 Db 344 FTN-----WLRNGY-----RGFLGFGAASNDTCNOA-VANMLTFVXNNADY----- 385  
 QY 315 QWTFWSWNPDSGDTGILK--DDWQTVTDVKGYLAP 349  
 Db 386 -WTGMAWAGGPPWGGYMYSTPEPSNGVDPKQMSVLAP 421  
 RESULT 10  
 GUN2\_RALSO STANDARD; PRT; 426 AA.  
 ID GUN2\_RALSO  
 AC P17974;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN EGL.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;

28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase B precursor (EC 3.2.1.4) (EGB) (Endo-1,4-beta-glucanase)  
DE (Cellulase B).  
GN CELS.  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
RN NCBI\_TaxID=1515;  
RX STRAIN=FROM N.A.  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB 10882;  
RX MEDLINE=86148508; PubMed=3453102;  
RA Grepinet O., Beguin P.;  
RT "Sequence of the cellulase gene of Clostridium thermocellum coding  
for endoglucanase B";  
RL Nucleic Acids Res. 14:1791-1799 (1996).  
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-  
GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
GLUCANS.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS  
WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN  
MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
hydrolases).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X03592; CAA27266.1; -.  
DR PIR; A23512; CZCLM.  
DR HSP; P54583; 1ECC.  
DR InterPro; IPR002105; Dockerin\_1.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00150; cellulase; 1.  
DR Pfam; PF00404; Dockerin\_1; 2.  
DR PROSITE; PS00018; EF HAND; UNKNOWN 1.  
DR PROSITE; PS00448; CLOS\_CELLULOSE\_RPT; 2.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
FT SIGNAL 1 27 OR 31.  
FT CHAIN 28 563 ENDOGLUCANASE B.  
FT ACT\_SITE 204 204 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 363 363 NUCLEOPHILE (BY SIMILARITY).  
FT DOMAIN 502 557 2 X 24 AA APPROXIMATE REPEATS.  
FT REPEAT 502 526 1.  
FT REPEAT 534 556 2.  
SQ SEQUENCE 563 AA; 63929 MW; 866FE55704A1DE4B CRC64;  
Query Match 26.8%; Score 533; DB 1; Length 563;  
Best Local Similarity 31.0%; Pred. No. 1.1e-34;  
Matches 130; Conservative 58; Mismatches 158; Indels 74; Gaps 13;  
QY 1 AGGYY-----W-HTSREILDANNVPRVIRAGINWFGFTCTVTVVHGLWSRDRVSM 50  
DB 27 AEGSYADLAEPDDDLHVEGTNIVDKYGNKVTGTANWFGNCREMLDLSYHSDIADI 86  
QY 51 DQKSLGYNTIRLPYSDDL- --KPGTTPNSINFRQNDQLGTLTSLQVMDKIVAYAGOI 107  
DB 87 ELVADKGINVRPIATDLLYANSQGIYPSTSTSYNNPALAGNSYELFNWLENFKRV 146  
QY 108 GLRIILDRHPCDGSQ- ---ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPH 164  
DB 147 GIKVILDVHSPETDNGHNTPLWYNTTITEEIFKXAVVWVAERYKNDNTIIGFDLKNPH 206  
QY 165 -----DPACWCGGDSIDWRLAAERAGNVLNVNPNLLIFVEGVQSY----- 206

207 TNTGTMKIKQAISALWDDSNHPNNWKRVAETALAILLEVHFNVLIFVEGVEMYPKDGWIDD 266  
QY -----NGDSY--WAGNLOGAGQYFVVL-NVFNRLVYSAHDYATSVYPTWFSDP 253  
DB 267 ETFTSPWTGNDYIGNWGNLGRGVVDYFINLCKYQSQVYVPHDYGPVIVQDFWFKGD 326  
QY 254 TFPNN-----MPGIWNKNGYLFNQNIAPVWLGEFGTTTLOSTTDQWTLKTLVQVLRP 305  
DB 327 FITANDEQAKRILVEQCWRDWAYIMEEGISPLLLGEWCGMTEG--GHPLLDLNLKYLRC 384  
QY 306 TAQYGADES---FQWTFWSWNPDSGDTGILKDD-----WQTVTVVDKGYLAPI 350  
DB 385 MRDFILENKYKLHHTFWCINIDSADTGLTREDGTFPGGRDLKMN--DNKYDNYLYPV 442  
RESULT 8  
GUNC\_PSEFL STANDARD; PRT; 748 AA.  
ID GUNC\_PSEFL  
AC P27033;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endoglucanase C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase) (EGC) (Cellodextrinase C).  
GN CELC.  
OS Pseudomonas fluorescens.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=294;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-47.  
RC STRAIN=Sp. Cellulosa;  
RX MEDLINE=92061996; PubMed=1953673;  
RA Ferreira L.M.A., Hazlewood G.P., Barker P.J., Gilbert H.J.;  
RT "The cellodextrinase from Pseudomonas fluorescens subsp. cellulosa  
consists of multiple functional domains.";  
RL Biochem. J. 279:793-799 (1991).  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)  
domain.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
hydrolases).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X61299; CAA43597.1; -.  
DR PIR; S19652; S19652.  
DR HSP; P07986; 1EXG.  
DR InterPro; IPR001919; Bac\_celose-bind.  
DR InterPro; IPR009031; CBDX.  
DR InterPro; IPR008965; Cellul bind.  
DR InterPro; IPR002883; Dockerin CBD 5.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF02013; CBM\_10; 1.  
DR Pfam; PF00553; CBM\_2; 1.  
DR Pfam; PF00150; cellulase; 1.  
DR SMART; SM00637; CBD II; 1.  
DR SMART; SM00561; CBD BACTERIAL; 1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 38  
FT CHAIN 39 748 ENDOGLUCANASE C.  
FT DOMAIN 139 136 CELLULOSE-BINDING.  
FT DOMAIN 137 179 SER-RICH (LINKER).  
FT DOMAIN 228 280 SER-RICH (LINKER).

DR Pfam; PF00404; Dockerin\_1; 2.  
 DR PROSITE; PS00018; EF\_HAND; UNKNOWN 2.  
 DR PROSITE; PS00448; GLOS\_CELLULOSOME\_RPT; 2.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 566  
 FT ACT SITE 226 226  
 FT ACT SITE 381 381  
 FT DOMAIN 503 549  
 FT REPEAT 503 526  
 FT REPEAT 536 549  
 SQ SEQUENCE 566 AA; 63199 MW; 2C9DE1AD87C3178 CRC64;

Query Match 30.6%; Score 608.5; DB 1; Length 566;  
 Best Local Similarity 34.5%; Pred. No. 1.2e-40;  
 Matches 140; Conservative 54; Mismatches 143; Indels 69; Gaps 15;

QY 6 W-HTSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDYRSMDOIKSLGVNIRLP 64  
 DB 44 WLCKGNKIYDMYGNEVWLTGANWFGFENSCNCPHGAW-YDVKTILTSIADRGINLIRLP 102  
 QY 65 YSDDIL-----KFG-----TMSNSINFRMNQDLOG-----LTSQVMDKIVAYAGQI 107  
 DB 103 ISTELLYSMWIKGNPVSSTASNNPPVHVVPDFYDPETDDVKNSEIFDIIMGYCKEL 162  
 QY 108 GLRIILDRHDPDC--SQGS-ALWY---TSS---VSEATWISDLOALAQRYKGNPTVVGF 158  
 DB 163 GIKWIMDIHSDANNSGHNYELWYKGTSTCGVVTTRKWDITLWLDADRYKNDITIAFD 222  
 QY 159 LHNPHDP-----ACWCGDPSIDWRLAARAGNAVSNPNLLIFVEGVQSY- 206  
 DB 223 LKNEPHGKRGYTAEPVKLLAKWDNSTDENWYKAETCAKALLENPKVLIVIEGVQYP 282  
 QY 207 -----NGD-----SYWGNLQAGQYPPVVLNPN-RLVYSAHDYATSV 244  
 DB 283 KTEKGYTYDPTDIWGATGASPMYSWNGNLRGVKDYDIPDLGPNLSQIVYSPHYGFSV 342  
 QY 245 YPQTFW-SDPTFPNNMPCINWKNWGLFNQNIAPVWLGEFTTLQSTTDQTLKTLVOYL 303  
 DB 343 YAQPMFEKDFVTQTLDDYDVTWAYIHQGIAGIILIGEWGHHGMDGKNQKWTLLRDYI 402  
 QY 304 RPTAQYAGDSFQWTFWSPNPSGDTGGILKDDQWDTVTKDGYLAP 349  
 DB 403 -----VQNRTHHTFWCINPNSGDTGGLGNDWSTWDEAKYALLKP 442

RESULT 6  
 GUND CELFI  
 ID GUND CELFI STANDARD; PRT; 747 AA.  
 AC P50400;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN CEND.  
 OS Cellulomonas fimi.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Micrococccineae; Cellulomonadaceae; Cellulomonas.  
 OX NCBI\_taxid=1708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93209933; PubMed=8458833;  
 RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.;  
 RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase  
 RT D (Cend), a family A beta-1,4-glucanase.";  
 RL J. Bacteriol. 175:1910-1918(1993).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- PATHWAY: Cellulose degradation.  
 CC -!- SIMILARITY: Contains 2 fibronectin type III domains.  
 CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)

CC domain  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L02544; AAA23089.1; -.  
 DR HSSP; P07986; 1EXG.  
 DR InterPro; IPR001919; Bac\_celose-bind.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR Pfam; PF00553; CSM 2; 1.  
 DR Pfam; PF00150; cellulase; 1.  
 DR Pfam; PF00041; fn3; 2.  
 DR SMART; SM00637; CBD\_II; 1.  
 DR SMART; SM0060; FN3; 2.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 KW SIGNAL 1 39  
 FT CHAIN 40 747  
 FT DOMAIN 448 542  
 FT DOMAIN 546 639  
 FT ACT SITE 208 208  
 FT ACT SITE 349 349  
 FT ACT SITE 349 349  
 SQ SEQUENCE 747 AA; 78936 MW; BD15473C9D8B42BD CRC64;

Query Match 27.7%; Score 549.5; DB 1; Length 747;  
 Best Local Similarity 34.5%; Pred. No. 7.8e-36;  
 Matches 130; Conservative 59; Mismatches 143; Indels 45; Gaps 14;

QY 1 AGGGYV-HTSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDYRSMDOIKSLGVN 59  
 DB 40 ATGDDWLHVEGTVIVDSTGKEALLSGVNWFGFENASERVHGLWSGNITQITQMAQRGIN 99  
 QY 60 TIRLPYSDDIL---KPTGM---PNSINFRMNQDLOGTLQVMDKIVAYAGQIGRLIILD 114  
 DB 100 VVRVPVSTQLLEWKAGTFLKPNVNTY--ANPELEGKNSLQIFEFYWLTLCKQYGIKVFEL 157  
 QY 115 RH--RDCSGQ--SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPH-----D 165  
 DB 158 VHSADNSCHVYNNWKKDITTEDVYEGWEWAATFKWDDDTIVGADIKNEPHGQTGST 217  
 QY 166 PACWCGDPSIDWRLAARAGNAVSNPNLLIFVEGVQSYN-----GDSY--W 212  
 DB 218 RAKWDGTTDKNFKHAETASKKILAINPNWLVFVEGVETPKPGVPWTSTGLTDYGTW 277  
 QY 213 WGNLQAGQYPPVVLNPN-RLVYSAHDYATSVYQTFW-SDPTFPNNMPCINWKNWGLV 270  
 DB 278 WGNLGRVDRHFDJLCAHQDQVLYPSHDYGLVFKQKFKQDFDKASLTADWGNWLF 337  
 QY 271 FNQNIAPVWLGEFTTL-QSTTDQW---LKTIVQVLRPTAQYAGDSFQWTFWSPNPSG 326  
 DB 338 HDEDIAFLIGEWGGRIGDPRQDKWMAALRDLVAER-----LSQTFWVLNPN 388  
 QY 327 DTGGLKDDQWDTVTK 343  
 DB 389 DTGGLLDDDKTWDEVK 405

RESULT 7  
 GUND CLOTM  
 ID GUND CLOTM STANDARD; PRT; 563 AA.  
 AC P04956;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)







PRO/SER/THR-RICH (LINKER).  
CELLULOSE-BINDING (BY SIMILARITY).  
PROTON DONOR.  
NUCLEOPHILE.

401 461  
462 562  
203 203  
323 323  
75 161  
209 212  
48 50  
51 52  
53 55  
57 58  
61 62  
65 68  
72 72  
73 73  
75 76  
80 84  
87 96  
97 98  
101 107  
108 111  
113 114  
127 129  
130 131  
134 147  
148 149  
151 158  
160 160  
162 162  
163 163  
175 188  
189 191  
193 194  
195 199  
206 207  
209 209  
215 217  
218 218  
219 230  
231 233  
235 236  
242 242  
244 244  
246 247  
248 249  
250 251  
254 255  
257 257  
259 263  
274 279  
283 285  
289 292  
294 299  
300 307  
308 308  
309 313  
314 315  
319 323  
331 343  
344 344  
347 350  
351 352  
359 359  
367 369  
373 373  
375 376  
380 380  
382 387  
388 388  
389 391  
SEQUENCE 562 AA; 60747 MW; 84E6256406A35041 CRC64;  
Query Match 99.6%; Score 1979; DB 1; Length 562;  
Best Local Similarity 99.7%; Pred. No. 4.9e-149;

Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AGGYWHTSGREILDANNVVRVIRAGINWFGFCTCNVYVHGLWSDYRSMLDQIKSLGYNT 60  
DB 42 AGGYWHTSGREILDANNVVRVIRAGINWFGFCTCNVYVHGLWSDYRSMLDQIKSLGYNT 101  
QY 61 IRLPYSDIILKPGTMPNSINFRONQDLQGLTSLQVMDKIVAYAGQIGLRIIILDRHRPDC 120  
DB 102 IRLPYSDIILKPGTMPNSINFRONQDLQGLTSLQVMDKIVAYAGQIGLRIIILDRHRPDC 161  
QY 121 SGQSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 162 SGQSALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 221  
QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOQAGQYPPVNLNPNRLVYSADY 240  
DB 222 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOQAGQYPPVNLNPNRLVYSADY 281  
QY 241 ATSVYPTWFSDDPTFNNMPCINWKNWGYLFNQNIAPVWLGEFGTTLQSTTDDOTLWLTIV 300  
DB 282 ATSVYPTWFSDDPTFNNMPCINWKNWGYLFNQNIAPVWLGEFGTTLQSTTDDOTLWLTIV 341  
QY 301 QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDWQIVDTVKGYLAPIKSSIFDPV 358  
DB 342 QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDWQIVDTVKGYLAPIKSSIFDPV 399

RESULT 2

GUN\_PAPO  
ID\_GUN\_PAPO STANDARD; PRT; 397 AA.  
AC P23548;  
DT 01-NOV-1991 (Rel. 20, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase (EC 3.2.1.4) (endo-1,4-beta-glucanase) (Cellulase).  
OS Paenibacillus polymyxa (Bacillus polymyxa).  
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxID=1406;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90170877; PubMed=2307659;  
RA Baird S.D., Johnson D.A., Seligy V.L.;  
RT "Molecular cloning, expression, and characterization of  
endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus  
circulans";  
RL J. Bacteriol. 172:1576-1586(1990).  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
hydrolases).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; M33791; AAA22631.1; --  
PIR; A35136; A35136.  
HSP; P54583; LECE.  
InterPro; IPR001547; Glyco\_hydro\_5.  
Fam; PF00150; cellulase; 1.  
PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
Cellulose degradation; Hydrolase; Glycosidase.  
FT ACT SITE 194 194 PROTON DONOR (BY SIMILARITY).  
FT ACT SITE 317 317 NUCLEOPHILE (BY SIMILARITY).  
SQ SEQUENCE 397 AA; 44357 MW; B9C2E802C04FOA2A CRC64;

Query Match 52.8%; Score 1048; DB 1; Length 397;  
Best Local Similarity 54.2%; Pred. No. 1.3e-75;  
Matches 193; Conservative 45; Mismatches 106; Indels 12; Gaps 4;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:59 ; Search time 8.33333 Seconds  
(without alignments)  
2236.934 Million cell updates/sec

Title: US-09-997-504A-14

Perfect score: 1986

Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSIIDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1979	99.6	562	1 GUN1_ACICE	P54583 acidothermu
2	1048	52.8	397	1 GUN_PAEPO	P23548 paenibacill
3	803.5	40.5	484	1 GUN4_XANCP	P19487 xanthomonas
4	626.5	31.5	1039	1 GUNB_CALSA	P10474 c endogluc
5	608.5	30.6	566	1 GUNG_CLOTM	Q05332 clostridium
6	549.5	27.7	747	1 GUNB_CLEFI	P50400 cellulomona
7	533	26.8	563	1 GUNB_CLOTH	P04956 clostridium
8	298.5	15.0	748	1 GUNC_PSEFL	P27033 pseudomonas
9	184	9.3	424	1 GUN1_RALSO	P58599 ralstonia s
10	166	8.4	426	1 GUN2_RALSO	P17974 ralstonia s
11	156.5	7.9	814	1 GUNE_CLOTH	P10477 clostridium
12	146.5	7.4	406	1 GUN1_RUMAL	P16216 ruminococcu
13	143.5	7.2	584	1 GUNB_CLOCE	P25472 clostridium
14	137	6.9	900	1 GUNH_CLOCE	P16218 clostridium
15	134	6.7	409	1 GUNB_RUMAL	P23661 ruminococcu
16	132	6.6	438	1 EXG_CANAL	P29717 candida alb
17	128.5	6.5	388	1 GUN3_HUMIN	Q12624 humicola in
18	128	6.4	448	1 GUN_CLOSA	P15704 clostridium
19	124	6.2	440	1 GUNB_CLOCL	P28621 clostridium
20	123	6.2	825	1 GUN3_BACSA	P19570 bacillus sp
21	122.5	6.2	517	1 GUN4_CLOLO	P54937 clostridium
22	120	6.0	336	1 GUNA_RUMFL	P16169 ruminococcu
23	118	5.9	501	1 YBQ6_YEAST	P18081 saccharomyc
24	117	5.9	445	1 SPRI1_YEAST	P32603 saccharomyc
25	116	5.8	562	1 EXG2_YEAST	P52911 saccharomyc
26	114.5	5.8	658	1 GUN3_FIRSU	P14250 fibrobacter
27	112	5.6	764	1 YIS7_YEAST	P40566 saccharomyc
28	110.5	5.6	505	1 GUNV_ERWCA	Q47096 erwinia car
29	110.5	5.6	532	1 HEXA_DICDI	P13723 dictyostell
30	109.5	5.5	418	1 GUN2_TRIRE	P07982 trichoderma
31	109.5	5.5	429	1 GUNA_BUTFI	P22541 butyrivibri
32	108.5	5.5	459	1 GUNA_STRLI	P27035 streptomyc
33	108.5	5.5	466	1 GUN5_THEFU	Q01786 thermomonos

34	108	5.4	312	1 GUN4_RUMAL	Q07940 ruminococcu
35	107.5	5.4	475	1 GUNA_CLOCE	P17901 clostridium
36	106.5	5.4	547	1 GUN1_BUTFI	P20847 butyrivibri
37	106	5.3	444	1 GUNN_ERWCA	Q59394 erwinia car
38	104.5	5.3	364	1 GUNA_RUMAL	P23660 ruminococcu
39	104.5	5.3	504	1 GUNW_ERWCA	Q59395 erwinia car
40	104	5.2	383	1 MANA_STRLI	P51529 streptomyc
41	101	5.1	515	1 GUNB_CLOCL	P28623 clostridium
42	100.5	5.1	343	1 GUNC_CLOTH	P07985 clostridium
43	100.5	5.1	499	1 GUN2_BACSU	P10475 bacillus su
44	100.5	5.1	800	1 GUN_BACSI	P06564 bacillus sp
45	99.5	5.0	499	1 GUN3_BACSU	P23549 bacillus.su

RESULT 1

ALIGNMENTS

100% ID GUN1\_ACICE STANDARD; PRT; 562 AA.

AC P54583;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)  
DE (Cellulase E1) (Endocellulase E1).  
OS Acidothermus cellulolyticus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Frankineae; Acidothermaceae; Acidothermus.  
OX NCBI\_TaxID=28049;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 43068 / 11B;  
RA Laymon R.A., Himmel M.E., Thomas S.R.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.  
RX MEDLINE=96346058; PubMed=8718854;  
RA Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kapus P.A.;  
RT "Crystal structure of thermostable family 5 endocellulase E1 from  
RL Acidothermus cellulolyticus in complex with cellobiose.";  
RL Biochemistry 35:10648-10660(1996).  
CC -!- FUNCTION: Has a very high specific activity on  
CC carboxymethylcellulose.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- MISCELLANEOUS: Thermostable enzyme with an optimal temperature of  
CC 81 degrees Celsius.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
CC hydrolases).  
-----  
CC This SWISS-PROT entry is copy-right. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: U33212; AAA75477.1; --  
CC PDB: 1ECE; 14-OCT-96.  
CC PDB: 1COD; 23-JUN-99.  
CC InterPro: IPR001919; Bac\_celose-bind.  
CC InterPro: IPR008965; Cellul\_bind.  
CC InterPro: IPR001547; Glyco\_hydro\_5.  
CC Pfam: PF00553; CBM 2; 1.  
CC Pfam: PF00150; cellulase; 1.  
CC SMART: SM00637; CBD II; 1.  
CC PROSITE: PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
CC Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.  
CC SIGNAL 1 41  
CC CHAIN 42 562 ENDOGLUCANASE E1.  
CC DOMAIN 42 400 CATALYTIC.

This Page Blank (uspto)

QY 68 DILKPGTWPNSINFRQ-----MNQDLQ-----LTSLOVMDKIVAYAGQIGLRI 111  
 DB 66 DLTNETLIANNVTYRQSFQSLGNDIVGFRTNNPSIIDPLIEAYKTVVTTLGNNDVMV 125  
 QY 112 ILDRH--RPD--CSGQ--SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHD 165  
 DB 126 ILDNHLLTKPGWCCANDGCGFFGDFPTVWVAALKMAATFNGSVNVGMSLRNELRG 185  
 QY 166 PACWCGDPSIDWRLAERAGNAVLNPNLLIFVEGVQSYNGDSYWWGGLQAGQYPV 225  
 DB 186 PK-----QNVNDFKYMGGAGAEVHSAANNKVLVILSLG-SPDAD-----LSFVRSRPV 232  
 QY 226 VLNVNRLVYSAHDYATSVYPTWFS-----PTFPNNMPG-IWNK--NWGYLFNQIA 276  
 DB 233 KLSFTGKLVPELHWYS-----FSDGNSWAANNPNDICORVLNRIKNGGGYLLNQGF- 283  
 QY 277 PWLGEFOTTLQ--STDTQWLKTLVQYLRPTAQYGADSFQWTFW-----SWNPDSDGTG- 329  
 DB 284 PLFLSEFGIDRGVNTDNRVFGCL-----TGWAENDVDWLSLWALTGSYTLRQGVGM 337  
 QY 330 ----GILKDDWQTV 339  
 DB 338 NEYGVLDSDWISV 351

RESULT 15  
 T51476  
 hypothetical protein K3M16\_70 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
 C:Accession: T51476  
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Men  
 submitted to the Protein Sequence Database, August 2000  
 A:Reference number: Z25394  
 A:Accession: T51476  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-526 <SAP>  
 A:Cross-references: EMBL:AL391150  
 A:Experimental source: cultivar Columbia; BAC clone K3M16  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 135/3; 218/2; 327/2; 393/1  
 A>Note: K3M16\_70

Query Match 10.6%; Score 210; DB 2; Length 526;  
 Best Local Similarity 22.6%; Pred. No. 1e-08;  
 Matches 84; Conservative 64; Mismatches 159; Indels 64; Gaps 14;  
 QY 8 TSREILDANNVPRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNITIRLPYSD 67  
 DB 30 TKSRWIVNKGHRVRLACANW-PSHLKPVVAEGLSSQPMDSISKIKDMGFNCVRLTWPL 88  
 QY 68 DILKPGTWPNSINFRQ-----MNQDLQGLTS-----LQVMDKIVAYAGQIGLRI 111  
 DB 89 ELMINDTLAFNVTVKQSFERYVGLDHELGQIYTHNPYIVNTPLINVFQAVVYSLGRHDVMV 148  
 QY 112 ILDRHP-----DCSGQ--SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHD 165  
 DB 149 ILDNHKTVPGCCNDPDPAFFGDPKFPNPLWMLGLKKVATI FMNVKNVYVGMSLRNL- 207  
 QY 166 PACWCGDPSIDWRLAERAGNAVLNPNLLIFVEGVQSYNGDSYWWGGLQAGQYPV 225  
 DB 208 ----GYNHTSKDWKYMKQGAEAHTNPVNLVILSGN-----FDADLSFLKDRPV 255  
 QY 226 VLNVNRLVYSAHDYATSVYPTWFSPT--FPNNMPGIWNKNKGYLFNQIAVPWLGEF 283  
 DB 256 NLSFKKLVLELHWYSFTDGTGQWKNHNVNDFCSQMFSEKERRTGGFVLDQGF-PLFLSEF 314  
 QY 284 GT-----TLQSTTTDTQWLKTLVQYLRPTAQYAD-SFQWTFWS-----WNPDSGDTG--- 329  
 DB 315 QTDQGGDLEGN-----RYNMCMLAAEAKDLDAWAVATGVYFPEGKRGVVEA 364

QY 330 -GILKDDWQTV 339  
 DB 365 YGMLDANWENV 375

Search completed: August 2, 2004, 16:37:40  
 Job time : 13.3333 secs

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Praga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig Chado, M.A.; Madeira, N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.A.; Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak A.; Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF0810

Query Match 16.3%; Score 324; DB 2; Length 565; Best Local Similarity 29.6%; Pred. No. 1.2e-17; Matches 103; Conservative 54; Mismatches 135; Indels 56; Gaps 12;

QY 12 EILDANNV-----PVRIAGINWFFETCNVYVHGLWSRDSYRMLDQIKSLGYNTRL 63

Db 4 ETMDAQNVSDFHVEPLH--GVNWGLETAQHLLQGLDQRPWREVLQDQWQSIGINAIRL 61

QY 64 PYSDDILKPTMPNSINFRQMODLOGLSLQWMDKIVAYAGIQLRIILDRHPC-SG 122

Db 62 PLCSDTLH-GTMTNLDLVR-NPDLKGRALQIADAIIDEAGKGRMVLAYHVECPD 119

QY 123 QSALWTSVSEATWISDLOALQRYKGNPTVV-GFDLHNEHPDACWCGDPSIDWRLA 181

Db 120 GNPLLSVDESEHQSVDQFITSFRAQKQVGVGVLDADMANHRPFQSGSSTPDNREV 179

QY 182 AERAGNAVLSVNPNLIFVEGV---QSYNGDSYVWGNLQAGQVYVNLVNPNR-LVYS 236

Db 180 VERAAAAILAMTEWLVGVQVGLNPPCLDASAPISDDNIQ--SQHCVPILRIPARNLLM 237

QY 237 AHYATSVYPTWFSDEPTFFNNPFGIWNKVGWLFNQIAPVWLGFG--TTLQSTTQ 293

Db 238 PHFAGTDIDTE-----AALGAFSGQTVLPSLDA 267

QY 294 TWLKTLYQLRPTAQYAGDSFQWTFWSNPDSDGTGILKDDQWQTVDT 341

Db 268 TDAEQLAHRDALLAFGR--QGYGSMWTSQAQWPFGLMNDGRTPT 313

RESULT 13

S19652

C:Species: Pseudomonas fluorescens

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 15-Oct-1999

C;Accession: S19652; S18849

R;Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J. Biochem. J. 279, 793-799, 1991

A;Title: The cellobiohydrolase from Pseudomonas fluorescens subsp. cellulosa consists of m

A;Reference number: S19652; MUID:92061996; PMID:1953673

A;Accession: S19652

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-748 <FER>

R;Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J. submitted to the EMBL Data Library, August 1991

A;Description: The cellobiohydrolase from Pseudomonas fluorescens subsp cellulosa consist

A;Reference number: S16849

A;Accession: S16849

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-262, 'C', 264-291, 'K', 293-748 <FE2>

A;Cross-references: EMBL:X61299; NID:945501; PID:945502

C;Genetics:

A;Gene: celC

C;Superfamily: bacterial cellulose-binding domain homology; glycosidase GNGW domain hom

F:38-134/Domain: bacterial cellulose-binding domain homology <BCB>

F:183-220/Domain: glycosidase GNGW domain homology <GWG>

F:39-133/Disulfide bonds: #status predicted

Query Match 15.0%; Score 298.5; DB 2; Length 748; Best Local Similarity 23.4%; Pred. No. 1.8e-15; Matches 111; Conservative 61; Mismatches 138; Indels 165; Gaps 23;

QY 7 HTSGREILDANNVPVRIAGINWFFETCN-----YVHGLW-----S 43

Db 290 NTQGNLTKDQGLLPARGC--NWFELEGRHEPNSDADNPSCGAPMELYAGNMWVYNSQSG 347

QY 44 RYRSMLDQIKSLGYNTRLRIPYSDDIL-----KGTWPNISNFRQMODLOGLSLQWMDK 99

Db 348 RTIQQTWTELKQQTWLRPIAPQTLDANDPQSRSENLKHKHSIRO-----NARQALD 403

QY 100 IVAYAGIQLRIILDRH-----RPDCSGQSALWY-----TSSVS 133

Db 404 FIKLADQNDIQIFDIHSCSNYVWGRAGRLDAPPYVDANRVGYDFTRREYSCSATNPS 463

QY 134 EAT-----WISDLOAL--QRYKGNPTVVGFDLHNEHPDACWCGDPSIDWRLA 182

Db 464 SVTRIHAHYDKQKWLANLREIAGLSAKLGVSNLIGIDVFNEPYD-YTWA-----EWKGV 516

QY 183 ERAGNAVLSVNPNLIFVEGVQSYNGDSY-----WVWGNLQAGQY 223

Db 517 EEAQAINVNPNLIIVEGI--SANANTQDGTPTDTSVVPVPHGSTDLNPNNGENLYEAGN 575

QY 224 PVNLNVP-NRLVYSAHYATSVYPTWFSDEPT-----FPNNWPG 261

Db 576 PP--NIPKDRLLFSPTVGPVSFVQRFMDPAQTECAGLEGEAAQARCRIVINPTVLEQ 633

QY 262 IWNKNGVLFNQIAPVWLGFG-----TTLQSTTQDTWLTLYQL 303

Db 634 GWEEHGVLRGLGV--ILIEFGGNWDPGAKSSQADNASHITTVDDQWQQAASYSF 692

QY 304 RPTAQYAGDSFQWTFWSNPDSDGTGILKDDQWQTVDTVKDGYLAPIKSSIFDVS 358

Db 693 K---RKGINA---CVWSMNPESADTMG-----YLTTP-----WDPV 722

RESULT 14

E86265

protein F3F19.15 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: E86265

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H. A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I. ket, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W. A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis. A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: E86265

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-522 <STO>

A;Cross-references: GB:AE005172; NID:94850396; PID:AAD31066.1; GSPDB:GN00141

C;Genetics:

A;Gene: F3F19.15

A;Map position: 1

Query Match 13.0%; Score 257.5; DB 2; Length 522; Best Local Similarity 25.1%; Pred. No. 1.9e-12; Matches 94; Conservative 59; Mismatches 150; Indels 71; Gaps 18;

QY 8 TSGREILDANNVPVRIAGINWFFETCNVYVHGLMSRDSYRMLDQIKSLGYNTRLPSD 67

Db 7 TSSRWIVDNGRLVKLVCAW--PSHLQPVVAEGLSKQPVDAVAKKIVEMGFCVRLTWPL 65

Db 343 YAQPPEKDFWTQTLDDYDWTWYTHDQGIAPILIGEWGHHMDGGKQKQKMTLLRDYI 402  
QY 304 RPTAQYAGDSFOWTFSWNPDSGDTGGILKDDMQTVDVTKDGYLAP 349  
Db 403 -----VQNRHHTFCINPNSGDTGGILGNDWSTWDEAKYALLKP 442

RESULT 10  
B47093  
cellulase (EC 3.2.1.4) Cend - Cellulomonas fimi  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Cellulomonas fimi  
C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 06-Dec-1996  
C:Accession: B47093  
R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.  
J. Bacteriol. 175, 1910-1918, 1993  
A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (Cend),  
F:644-745/Domain: bacterial cellulose-binding domain homology <BCB>  
A:Reference number: A47093; MUID:93209933; PMID:8458833  
A:Accession: B47093  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-747 <MEI>  
A:Experimental source: ATCC 484  
A>Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:128120, NCBIP:128122)  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
A:Pathway: cellulose degradation  
C:Superfamily: bacterial cellulose-binding domain homology  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:644-745/Domain: bacterial cellulose-binding domain homology <BCB>  
F:645-744/Disulfide bonds: #status predicted

Query Match 27.7%; Score 549.5; DB 2; Length 747;  
Best Local Similarity 34.5%; Pred. No. 3.6e-35;  
Matches 130; Conservative 59; Mismatches 143; Indels 45; Gaps 14;  
QY 1 AGGGYW-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYN 59  
Db 40 ATGDDWLHVEGNTIVDSTGKALISGVNWFNFASERVFHGLNSGNTQITQQAQGIN 99  
QY 60 TIRLPYSDDL---KPGTM--PNSINFRQMDQLGTSIQVMDKIVAYAGQIGIRILD 114  
Db 100 VVRVPSVQQLLEWKAQTFKPNVNTY--ANPELEGKNSLQIFBYWLTLCQYGIKVFLD 157  
QY 115 RH--RPPCSGQ-SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPH---D 165  
Db 158 VHSRAEDNSGHVYNNWKGDIITTEDVYEGWEAAATRNKDDTIIVGADIKNEPHGTQSGTE 217  
QY 166 PACMGCGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYN-----GDSY--W 212  
Db 218 RAKWDGTTDKNFKHFATASKKILAINPNLVFVEGVIEYKPGVPTWSTGLTDYVGTW 277  
QY 213 WGNHLOGAGQYPVVLN--PNRLVYSAHDYATSVYPTWF--SDPTFPNPNKGNKGYL 270  
Db 278 WGNLGRVDRPHDILGAHQDQLVYSPHDYGPLVFDQKWFQKDFKASLTADVWGPWLFI 337  
QY 271 FNQNIAPVNLGDFGFTL--QSTTDQW---LKTLYQYLRPTAQYAGDSFQWTFWSNPDGS 326  
Db 338 HDEIAPLLIGEWGRLGQDPRQDKWAAALDLVAERR-----LSQTFWLVNENSG 388  
QY 327 DTGILLKDDWQTVTVK 343  
Db 389 DTGILLDDWKTWDEVK 405

RESULT 11  
CZCLBM  
cellulase (EC 3.2.1.4) B precursor - Clostridium thermocellum  
N:Alternate names: endo-1,4-beta-glucanase B precursor  
C:Species: Clostridium thermocellum  
C>Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 18-Jun-1999

C:Accession: A23512  
R:Grepinet, O.; Beguin, P.  
Nucleic Acids Res. 14, 1791-1799, 1986  
A:Title: Sequence of the cellulase gene of Clostridium thermocellum coding for endoglucan  
A:Reference number: A23512; MUID:86148508; PMID:3453102  
A:Accession: A23512  
A:Molecule type: DNA  
A:Residues: 1-563 <GRE>  
A:Cross-references: GB:X03592; NID:940668; PIDN:CAA27266.1; PID:940669  
C:Comment: This secretory enzyme is part of a highly active and thermostable cellulase co  
C:Genetics:  
A:Gene: celB  
C:Function:  
A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans  
A:Pathway: cellulose degradation  
C:Superfamily: cellulase B; Clostridium cellulase repeat homology  
C:Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharide de  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:26-563/Product: cellulase B #status predicted <MAT>  
F:502-525/Domain: Clostridium cellulase repeat homology <CCR1>  
F:534-557/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 26.8%; Score 533; DB 1; Length 563;  
Best Local Similarity 31.0%; Pred. No. 4.9e-34;  
Matches 130; Conservative 58; Mismatches 158; Indels 74; Gaps 13;  
QY 1 AGGGY-----W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSML 50  
Db 27 AEGSYADLAEPDDDLHVEGTVIVDKYGNKVIWITGANWFGFNCRERMLDSTHSDIADI 86  
QY 51 DQIKSLGYNTIRLPYSDDL---KPGTMFNSINFRQMDQLGTSIQVMDKIVAYAGOI 107  
Db 87 ELVADKGINVVRMFIATDLLYAWSQGIYPPSTDTSNNPALAGLSYELENFMLENFKRV 146  
QY 108 GURIIIDRHRPCSGQS---ALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPH 164  
Db 147 GIKVILDVHSPETDNGHNYPLWYNTITTEEFKXAWVVAERYKNDTDITIGFDLHNEPH 206  
QY 165 -----DPACMGCGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYN----- 206  
Db 207 TNGTGWKIAQAQSAIMDDSNHFNKVAEETALALEVHPNVLIIVEGVEMYPKDGIMDD 266  
QY 207 -----NGDSY--WVGMLQAGQYPVVL--VNPRLVYSAHDYATSVYPTWFSDDP 253  
Db 267 ETPDTSFMTGNNDYGNWVGNGLRGVKDYPINLGKYQSOLVYSPHDYGIPIVYEQDMFKGD 326  
QY 254 TPENN-----MPGINKNKGVLNPNIAIPWILGERGTTIQTSTTDQWTKLTVQYLRP 305  
Db 327 FITANDEQAKRILYEQCWEDNWAYINEBGISPLLQEWGCGWTEG--GHPLLDNLNLYLR 384  
QY 306 TAYQYAGDS---FQWTFWSNPNPDGDTGGILKDD-----WQTVDTVTKDGYLAPI 350  
Db 385 MRDFILEYKYLHHTFPCINISADTGGLFTRDEGTPPGGRDLKWN--DNKYDNYLYPV 442

RESULT 12  
B82761  
extracellular endoglucanase precursor XF0810 [imported] - Xylella fastidiosa (strain 9a50  
C:Species: Xylella fastidiosa  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: B82761  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: B82761  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-565 <SIM>  
A:Cross-references: GB:AE003920; GB:AE003849; NID:g9105697; PIDN:AAF03620.1; GSPDB:GN0012  
A:Experimental source: strain 9a50  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al



F/20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

```
Query Match      31.5%; Score 626.5; DB 2; Length 915;
Best Local Similarity 35.5%; Pred. No. 4.3e-41;
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY      6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTIRLP 64
DB      505 WLYVSGNKIVDKGRFWLTGTINWFGYNTGTNVFDGWSCLNKTDLAEIANRGFNLLRVP 564

QY      65 YSDIL- --KPGTMPN- SINFRQMNODLOGLTSLQVMDKIVAYAGQIGLRIILDRH- --RP 118
DB      565 ISABELLANSOGIYFPKNINY- YNPELEGKNSLEVDIVVOTCKEVGLKIMLDIHSIKT 623

QY      119 DCSGO- SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP- -----ACWG 170
DB      624 DAMGHYIPVWYDEKFTPEDFYKACEWITNRYKNDDTIIAFDLKNEPHGKPMQDTTFKAWD 683

QY      171 CGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD- -----SYWGGNLL 217
DB      684 NSTDIINWKAETCAKRIILINPNLLIVIEGLEAYPKDDVTWTSKSSDDYISTWGGNLL 743

QY      218 QGAGQYPPVL- NVNPNLVYSAHDYATSVYPTQWFSPTFPNN- --MPGIWKNWGYLFNQN 274
DB      744 RGVKYPINLKYONKVVYSPHDYGPSVYQPMFY- --PGFTKESLLQDCWRPNWAYIMEEN 802

QY      275 IAPVWLGEFGTTLQSTTDQTLWKLTVQYLRPTAQYGADSFQWTFWSNPDSDGTGILKD 334
DB      803 IAPLLIGEWGHLGDADNEKMKYLRDYII- -----ENHIHTFWCFNANSBGDTGGLVGY 856

QY      335 DWQTVDTVKDGYLAP 349
DB      857 DFTTWDEKYSFLKP 871
```

## RESULT 8

```
S02711
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum
N/Alternate names: endo-1,4-beta-glucanase
N/Contains: cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)
C/Species: Caldocellum saccharolyticum
C/Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C/Accession: S02711
R/Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.
Nucleic Acids Res. 17, 439, 1989
A/Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo-
A/Reference number: S02711; MUID:89098398; PMID:2789517
A/Accession: S02711
A/Molecule type: DNA
A/Residues: 1-1039 <SAU>
A/Cross-references: EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646
C/Genetics:
A/Gene: celB
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
A/Pathway: cellulose degradation
C/Superfamily: Streptomyces endo-1,4-beta-xylanase A homology
C/Keywords: glycosidase; hydrolase; polysaccharide degradation
F/1-28/Domain: signal sequence #status predicted <SIG>
F/29-1039/Product: cellulase #status predicted <MAT>
F/72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>
```

```
Query Match      31.5%; Score 626.5; DB 2; Length 1039;
Best Local Similarity 35.5%; Pred. No. 5.1e-41;
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;
```

```
QY      6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTIRLP 64
DB      629 WLYVSGNKIVDKGRFWLTGTINWFGYNTGTNVFDGWSCLNKTDLAEIANRGFNLLRVP 688

QY      65 YSDIL- --KPGTMPN- SINFRQMNODLOGLTSLQVMDKIVAYAGQIGLRIILDRH- --RP 118
DB      689 ISABELLANSOGIYFPKNINY- YNPELEGKNSLEVDIVVOTCKEVGLKIMLDIHSIKT 747
```

```
QY      119 DCSGO- SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP- -----ACWG 170
DB      748 DAMGHYIPVWYDEKFTPEDFYKACEWITNRYKNDDTIIAFDLKNEPHGKPMQDTTFKAWD 807

QY      171 CGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD- -----SYWGGNLL 217
DB      808 NSTDIINWKAETCAKRIILINPNLLIVIEGLEAYPKDDVTWTSKSSDDYISTWGGNLL 867

QY      218 QGAGQYPPVL- NVNPNLVYSAHDYATSVYPTQWFSPTFPNN- --MPGIWKNWGYLFNQN 274
DB      868 RGVKYPINLKYONKVVYSPHDYGPSVYQPMFY- --PGFTKESLLQDCWRPNWAYIMEEN 926

QY      275 IAPVWLGEFGTTLQSTTDQTLWKLTVQYLRPTAQYGADSFQWTFWSNPDSDGTGILKD 334
DB      927 IAPLLIGEWGHLGDADNEKMKYLRDYII- -----ENHIHTFWCFNANSBGDTGGLVGY 980

QY      335 DWQTVDTVKDGYLAP 349
DB      981 DFTTWDEKYSFLKP 995
```

## RESULT 9

```
A40589
cellulase (EC 3.2.1.4) - Clostridium thermocellum
N/Alternate names: endo-1,4-beta-glucanase
C/Species: Clostridium thermocellum
C/Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jun-2000
C/Accession: A40589; S31381
R/Lemaire, M.; Seguin, F.
J. Bacteriol. 175, 3353-3360, 1993
A/Title: Nucleotide sequence of the celG gene of Clostridium thermocellum and characteri-
A/Reference number: A40589; MUID:93273701; PMID:8501039
C/Accession: A40589
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-566 <LEN>
A/Cross-references: GB:X69390; NID:g40677; PIDN:CAA49187.1; PID:g40678
C/Genetics:
A/Gene: celG
C/Function:
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel-
A/Pathway: cellulose degradation
C/Superfamily: cellulase B; Clostridium cellulase repeat homology
C/Keywords: glycosidase; hydrolase; polysaccharide degradation
F/503-526/Domain: Clostridium cellulase repeat homology <CCRI>
F/536-559/Domain: Clostridium cellulase repeat homology <CCR>
```

```
Query Match      30.6%; Score 608.5; DB 2; Length 566;
Best Local Similarity 34.5%; Pred. No. 5.8e-40;
Matches 140; Conservative 54; Mismatches 143; Indels 69; Gaps 15;
```

```
QY      6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLMDQIKSLGYNTIRLP 64
DB      44 WLHCKGNKIYDMYGNVWLGTGANWFGNCSNCFHGAW-YDKTKILTSTIADRGINLLRIP 102

QY      65 YSDIL- --KPG- -----TMPNSINFRQMNQDLOG- -----LTSQVMDKIVAYAGQI 107
DB      103 ISTEULYSWMLGKNPVSSTASNNPPIVYVNPDPYDPTDDVKNSEIFDIINGYCKEL 162

QY      108 GLRIILDRH- --SGQS-ALWY- --TSS- --VSEATWISDLQALAQRYKGNPTVVGFD 158
DB      163 GIKWMIDITHSPDANNSGHNYELWYKETSCTCGVVTNKMIDTLVWLADKYKNDDTIIAFD 222

QY      159 LNEPHDP- -----ACWGGCPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSY- 206
DB      223 LKNEPHGKRGYTAEPVKLLAKNDSTDENWKNYAAETCAKALLENPKLIVIEGVEQYP 282

QY      207 -----NGD- -----SYWGGNLOGAGQYPPVIVNPN-RLVYSAHDYATSV 244
DB      283 KTEKGYTDTPIWGTGADGASPTYSANWGNLGRVKDYPIDGLPLNSQIVTSPHDYGPSV 342

QY      245 YPQTFW-SDTFFPNMNPPIWKNWGYLFNQNIAPWLGEFGTTLQSTTDQTLWKLTVQYL 303
```

Db 26 YSISHGKVIDDKGN-QIOLKGSWFGFETTHNVVHGLWTRNWKKEFITQIQSMGLNAVRLP 84  
Qy 65 YSDILLKPGTGNPNINFRWNODLQGLTSLOVMDKIVAYAGQIGRLRIILDRHPDSCQS 124  
Db 85 FCPASLNSNTSPSSIDYNR-NPDQLGLSSLIQIMDKVVKLSDRGIYVMDHHTPDCAAIS 143  
Qy 125 ALWYTSVSEATWISDLOALAOYKGNPTVVGDLHNEPHDPAQWCGDPSIDWRLAAER 184  
Db 144 ELWYTDAYSEKQWIDDLRFVAHRVAVPVGVLGVLDVKNNEPHGRATWGTGDKPTDNTAVEH 203  
Qy 185 AGNAVLSPNPNLLIFVGVSYNGDS----YWGNGNLQAGQVPLVNLNP-NRLVVSAMD 239  
Db 204 AAAAILEAAPKWLIGVEGIGENPSCSSTIGHFWGENLEPMDCPTP-LKVPADHLLLIPIHV 261  
Qy 240 YATSVYPTWFSDDPTFNNMPPGIWKNKNGYLFNQNIAPVWLGEFGTTL--QSTTDQTLK 297  
Db 262 YGPDVYVQPYFNSPDPFNNWAAIWDKHFHFAKAGYA-MAIGFPGKYGEGDPRDVAQN 320  
Qy 298 TLVYLRLPTAQYG-ADSFQWTFWSPNPDSDGTGGILKDDQTV 339  
Db 321 ALVDYL---ISIGVTDTF---YMSINANSSDTGGLFRDDNNHV 357  
RESULT 5  
E75142  
endoglucanase PAB0632 - Pyrococcus abyss (strain Orsay)  
C:Species: Pyrococcus abyss  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C:Accession: E75142  
R:anonymous, Genoscope  
submitted to the EMBL Data Library, July 1999  
A:Description: Pyrococcus abyss genome sequence: insights into archaeal chromosome stru  
A:Reference number: A75001  
A:Accession: E75142  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-514 <Kaw>  
A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PID:g545836  
A:Experimental source: strain Orsay  
C:Genetics:  
A:Gene: celB-like; PAB0632  
C:Superfamily: cellulase B; Clostridium cellulase repeat homology  
Query Match 34.7%; Score 688.5; DB 2; Length 514;  
Best Local Similarity 41.6%; Pred. No. 2.7e-46;  
Matches 148; Conservative 58; Mismatches 121; Indels 29; Gaps 10;  
Qy 7 HTSGREILDANNV-----PVRIAGINWFGFETCNVYVHGLWDRVRSMLDQIKSLGYNT 60  
Db 30 YTAENGILFVQNTVTGKPKLYLHGVSFGEFLKDHVYVGLDXNWKDILKDKVRLGFNA 89  
Qy 61 IRLPYSDDILKPGTMN--SINFRWNODLQGLTSLOVMDKIVAYAGQIGRLRIILDRHP 118  
Db 90 IRLPFCSESIRPDRPSPERIN--ELNPDNLKNTLSLEIMEKIIEYANSIGLYILLDYHRI 148  
Qy 119 DCSQSALWYTSVSEATWISDLOALAOYKGNPTVVGDLHNEPHDPAQWCGDPSIDW 178  
Db 149 GCEIEPLWYNTENISEQYIKDWIFLAKRPKYPNVIGAIKKEPHGEAGWGTGDER-DF 207  
Qy 179 RLAAERAGNAVLSNPNLLIFVEGVQ-----SYNGDSYV--MGNGNLQAGQVPLVNLNP 227  
Db 208 RLPAEKVGREILKVAHPHLIFVEGTQVTHVENIDELIEKKGWTFWGENLMGVKDYFVRL 267  
Qy 228 NVFNRLVYSAHDYATSVYPTWFSDDPTFNNMPPGIWKNKNGYLFNQNIAPVWLGEFGTTL 287  
Db 268 -PRGKVVYSPHYGFSYVMDYFKSPDPFNNMPLIETHEFGYLDLNTLV-LGEWGNV 325  
Qy 288 QSTTDQTLKTLVYLRLPTAQYGADSFQWTFWSPNPDSDGTGGILKDDQTVDTVK 343  
Db 326 EG-LDKVQWDAFVKWLKKKIY-----NFFVWCLNPSSGDTGGILFDDWKTAVNEK 375  
RESULT 6

E97012  
probable non-processive endoglucanase family 5, secreted, Cella homolog secreted, dockerir  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: E97012  
R:Nolling, J.; Bretton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: E97012  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-482 <Kur>  
A:Cross-references: GB:AE001437; PID:AAK78888.1; PID:g15023812; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0912  
Query Match 34.3%; Score 681; DB 2; Length 482;  
Best Local Similarity 37.6%; Pred. No. 9.5e-46;  
Matches 140; Conservative 71; Mismatches 127; Indels 34; Gaps 9;  
Qy 5 YHTSGREILDANNVPLRIAGINWFGFETCNVYVHGLWDRVRSMLDQIKSLGYNTIRLP 64  
Db 24 YLHSDGSKLLDDYGVQVMTGIANFLETPNYCFHGLWANRLDNLINIVADNGENTLRVP 83  
Qy 65 YSDILL---KPGT--MPNSINFRWNODLQGLTSLOVMDKIVAYAGQIGRLRIILDRHPD 119  
Db 84 LSELVQWNRQGVYPTPDSIN-DYISPELKQNSLIQLDDVIAVSKYGVKVMMLDHRIE 142  
Qy 120 CSQSALWYTSVSEATWISDLOALAOYKGNPTVVGDLHNEPHDPAQ-----WCGG 172  
Db 143 SGGQTATWYTSKYTTDDYKQWYLDYKNDTVAADIFNEPHGKAYRAETSAKNDT 202  
Qy 173 DPSIDWRLAAERAGNAVLSNPNLLIFVEGVQSY-----NGDSY---WCGNLQ 219  
Db 203 TDEDNRVYAEKGGKILIDIPKMLIVGVETYPKEGTAGSTNPDYVYGVWGNLVRG 262  
Qy 220 AGQYFVVL-NVENRLVYSAHDYATSVYPTWTF--SDPTFNNMPPGIWKNKNGYLFNQNIAP 277  
Db 263 VKDYFVDLAPYKNQVYVSPHDXPGVSDQTFWFDGDFTEQSLINDIWRPSWFYIOEKIAP 322  
Qy 278 VMLGFGTTLQSTTQTLKTLVQVLRPTAQYGADSFQWTFWSPNPDSDGTGGILKDDQ 337  
Db 323 LLIGSWGNGMDGKNEQWMTDMAKLI-----SDNMKNHTFWCLNANSNGTGGILEYDFK 376  
Qy 338 TVDTVVDKGYLAP 349  
Db 377 TIDTKLALVQ 388  
RESULT 7  
A43802  
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum sa  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Caldocellum saccharolyticum  
C:Date: 30-Jan-1993 #sequence\_revision 30-Sep-1993 #text\_change 10-Jul-1998  
C:Accession: A43802  
R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.  
Appl. Environ. Microbiol. 56, 3117-3124, 1990  
A:Title: celB, a gene coding for a bifunctional cell cellulase from the extreme thermophile "C  
A:Reference number: A43802; MUID:91136262; PMID:2126700  
A:Accession: A43802  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual trar  
A:Molecule type: DNA  
A:Residues: 1-915 <SAU>  
A:Cross-references: EMBL:X13602  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
A:Pathway: cellulose degradation  
C:Superfamily: Streptomyces endo-1,4-beta-xylosidase A homology  
C:Keywords: Glycosidase; hydrolase; polysaccharide degradation



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:59 ; Search time 12.3333 Seconds  
(without alignments)  
2792.154 Million cell updates/sec

Title: US-09-997-504A-14  
Perfect score: 1986  
Sequence: 1 AGGGYHTSGREILDANNVP.....VDTVKDGLAPIKSSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*  
1: Pirl1.\*  
2: Pirl2.\*  
3: Pirl3.\*  
4: Pirl4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1050	52.9	397	2 A35136	cellulase (EC 3.2.1.4)
2	803.5	40.5	493	1 JH0158	cellulase (EC 3.2.1.4)
3	762	38.4	458	2 E71059	probable endo-1,4-galactosidase
4	695.5	35.0	592	2 E82759	endo-1,4-beta-glucanase
5	688.5	34.7	514	2 E75142	endo-1,4-beta-glucanase
6	681	34.3	482	2 E97012	endo-1,4-beta-glucanase
7	626.5	31.5	1039	2 A43802	cellulase (EC 3.2.1.4)
8	626.5	31.5	1039	2 S02711	cellulase (EC 3.2.1.4)
9	608.5	30.6	566	2 A40589	cellulase (EC 3.2.1.4)
10	549.5	27.7	747	2 B47093	cellulase (EC 3.2.1.4)
11	533	26.8	563	1 C7CLBM	extracellular endo-1,4-beta-glucanase
12	324	16.3	565	2 B82761	cellulase (EC 3.2.1.4)
13	298.5	15.0	748	2 S19652	cellulase (EC 3.2.1.4)
14	257.5	13.0	522	2 E86255	cellulase (EC 3.2.1.4)
15	210	10.6	526	2 T51476	hypothetical prote
16	206.5	10.4	488	2 T51502	hypothetical prote
17	166	8.4	426	2 A42649	cellulase (EC 3.2.1.4)
18	156.5	7.9	814	1 C7CLBM	cellulase (EC 3.2.1.4)
19	154.5	7.8	516	2 E70134	mannan endo-1,4-beta-glucanase
20	154.5	7.8	722	2 H96986	endo-1,4-beta-glucanase
21	151.5	7.6	356	2 G82523	endo-1,4-beta-glucanase
22	147	7.4	1012	2 B97326	endo-1,4-beta-glucanase
23	146.5	7.4	329	2 C72216	endo-1,4-beta-glucanase
24	146.5	7.4	406	2 A43722	cellulase (EC 3.2.1.4)
25	143.5	7.2	584	2 J01229	cellulase (EC 3.2.1.4)
26	139.5	7.0	669	2 D72278	endo-1,4-beta-mannanase
27	137	6.9	900	2 JH0157	cellulase (EC 3.2.1.4)
28	136.5	6.9	500	2 S22458	cellulase (EC 3.2.1.4)
29	134	6.7	370	2 G97001	endo-1,4-beta-glucanase

#### ALIGNMENTS

##### RESULT 1

A35136  
cellulase (EC 3.2.1.4) - Bacillus polymyxa  
N: Alternate names: endo-1,4-beta-glucanase  
C: Species: Bacillus polymyxa  
C: Date: 03-Aug-1990 #sequence\_revision 03-Aug-1990 #text\_change 24-Sep-1998  
C: Accession: A35136  
R: Baird, S.D.; Johnson, D.A.; Seligy, V.L.  
J. Bacteriol. 172, 1576-1586, 1990  
A: Title: Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase  
A: Reference number: A35136; MUID: 90170977; PMID: 2307659  
A: Accession: A35136  
A: Status: preliminary  
A: Molecule type: DNA  
A: Residues: 1-397 <BAI>  
A: Cross-references: GB:M33791; GB:M33840  
C: Function:  
A: Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
A: Pathway: cellulose degradation  
C: Superfamily: Xanthomonas campestris cellulase  
C: Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 52.9%; Score 1050; DB 2; Length 397;

Best Local Similarity 54.2%; Pred. No. 7.9e-75; Mismatches 105; Indels 12; Gaps 4;

Matches 193; Conservative 46; Mismatches 105; Indels 12; Gaps 4;

QY 4 GYHHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMDDQIKSLGYNTIRL 63

DB 37 GYHTQGNKIVDESGKEAFAFNWFGLETPNYTLHGLWSRMDMDLPQVKKEGYNLRL 96

QY 64 PYSDILKPGTNPNINFRQMDLQGLTSLOVMDKIVAYAGQIGRLIILDRHRPDCSGQ 123

DB 97 PYSNOLFSSRPDSIDYHK-MPDLVGLNPQINDKLEKAGQIGIILDRHRPGSGGQ 155

QY 124 SALWTSSTVSEATWISDLOALAOYKGNPTVYVDFHNEPHDPACGCGDPSIDWRLAAE 183

DB 156 SELWTSQYPSERWISDWKMLADRYKNPTVIGADHNEPHGQASWGNGNASTDWRLAAQ 215

QY 184 RAGNAVLSVNPMLLIFVEGVQ---SYNGDSYWGNGNLAGQGYPVLLNVPRLVYSAHDY 240

DB 216 RAGNAVLSVNPMLLIFVEGVQ---SYNGDSYWGNGNLAGQGYPVLLNVPRLVYSAHDY 275

QY 241 ATSVTPQTFSPDPTFPNNMGGIWNKNWGLFNQNTAPVNLGEGF--TTLQSTDDQTLTKT 298

DB 276 GFGVSSQWPNFNDPAPPSNLPAIDWQTVGIVSKQNTAPVLVGEFGGNVDLSPEKQWNA 335

QY 299 LVQYLRLPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQVTDVTKDGYLAPIKSI 354

DB 336 LVHYI-----CANNLYFTYSLNPNDSGDTGGILLDDWTTWNRPKQDMLGRIMKPV 385

##### RESULT 2

*This Page Blank (uspto)*

```

; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9262
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9262

Query Match      33.4%; Score 662.5; DB 15; Length 421;
Best Local Similarity 39.9%; Pred. No. 2.2e-56;
Matches 141; Conservative 62; Mismatches 121; Indels 29; Gaps 12;

QY 5 YWHTSGREILDANNVVRVRIAGINWFGTETCNVYVHGLWSRDRSMLDQIKSLGYNTIRLP 64
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
26 YSISHGKVVDKGN QIQLRGVNWFETGTHVYVGLWANNWKEFITQLQGMGFNAIRLP 84
QY 65 YSDDILKPTMPSINFRQMNQDLOGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSQS 124
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
85 FCPANLNSNTSPSSIDYSR-NPDLQGLSSLQILDVVKVLSDRMYVLLDHRPDCSAIS 143
QY 125 ALWYTSSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWGGDPSIDWRLAER 184
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
144 ELWHTDSYSEKQWIDDLRFVARRVANVPGVIGLDVKNEPHGRATWGTGDFQTDWNTAVEH 203
QY 185 AGNAVLVSNPNLLIFVEGVQSYNGDS-----YWMGNLQAGQYFVVNLVNP-NELVY 235
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
204 AAAAILEAAPKWVIGIEI---GENPICSTIGHFWGENLEPMDCFP--LKVPADHLL 257
QY 236 SAHDYATSYVYQTFWESDPTFFNNMPTGNKNWGLFNQNIAPVWLGEFTL--QSTTDQ 293
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
258 MPHVYGPVYVYQYFNSPDPFNNMAAIWDKHFHFAKAGYA-MAIGFEGKYGEGDPRDI 316
QY 294 TWLKTLYOYLRPTAQYG-ADSFQWTFWSNPDSDGTGILKDDWQTV--DTVK 343
Db | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSGDTGGMVGNNDWTTPRDDKVK 363

```

Search completed: August 2, 2004, 16:40:50  
Job time : 42.3333 secs

[illegible]

RESULT 13  
 US-10-369-493-21618  
 ; Sequence 21618, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10(52052) B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 21618  
 ; LENGTH: 514  
 ; TYPE: PRT  
 ; ORGANISM: *Pyrococcus abyssi*  
 US-10-369-493-21618

Query Match	34.7%;	Score 688.5;	DB 15;	Length 514;
Best Local Similarity	41.6%;	Pred. No. 8.1e-59;		
Matches 148;	Conservative 58;	Mismatches 121;	Indels 29;	Gaps 10;
Qy	7	HTSGREILDANNV-----PVRAGINWFGTETCNVYVHGLWSRDYRSMLOQIKSLGYNT	60	
Db	30	YTAENGIIIVQNVITGKKPLYLHGVSNFGEELKSDHVVYGLDKENWKDILKDVKRLGNA	89	
Qy	61	IRLPYSDDILKPGWNPN--SINFQMNQDLOGLTSLQWMDKIYAVAGOIGRIILIDRRPP	118	
Db	90	IRLPFCSESIKPDTPSPERINY-ELNPDKNLAISLEIMEKIIIEYANSIGYIILLDYHRI	148	
Qy	119	DCSGOSALWYTSVSEATWISDQLALQARYKGNPTVVGFDLHNHPDHPACMGCDPSIDW	178	
Db	149	GCSEIEPLWYTENYSEEQYIKDWIFLAKRPGKYPNVICADIKNEPHGEAGMTGDER--DF	207	
Qy	179	RLAERAGNAVLVNPNLLIIFVEGVQ-----SYNGDSW--WGNLQAGQYFVVL	227	
Db	208	RLFAEKVGREILKVAPHWLIIFVEGQYTHVPNIDIEIKKGWMTFWGSLMGKVDKYPRL	267	
Qy	228	NVPNRLVYSADHYATSVYPTWFSDETPENNPNGIWNKNWGLYFNQNIAPVWLGEFGTTL	287	

```
Db      268 -PRGKVVSYPHYGSVTHMMYFKSPDFNNMPLIETHFGYLTDLNTLV-IGEGNGNY 325
          :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Qy      288 QSTTDTWLKTLTVQLRFTAQAGADSFOWTFWSWNPDSGTGGILKDDWOTVDVK 343
          :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
Db      326 EG-LDRVMQDAFVKVLKKIY----NFFVWCNLNPSGGTGGFLLDKMTVAWEK 375
          :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

RESULT 14
US-10-369-493-9473
; Sequence 9473, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTIVITY
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 473/4
; SEQ ID NO 9473
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9473
```

Query Match	33.6%	Score	667.5	DB	15	Length	369
Best Local Similarity	40.5%	Pred. No.	6e-57				
Matches	143	Conservative	58	Mismatches	123	Indels	29
Gaps	12						
QY	5	YWHTSGREILDANNVPVRIAGINWFGPETHYVYVHGLWSRDYRSMLDQIKSLGYNTIIRLP	64				
Db	26	YSISGHKVVDDKGN-QIQLRGVNWFGPETHVYVGLWARKWKEFITIQOGMGFNAIRLP	84				
QY	65	YSDDLKPGTWPNSINFRQWQDQLQGLTSLQVMDKIVAYAGCIGLRIILDRHRPDCSQS	124				
Db	85	FCPANLNSNTSPSIDYSR-NPDLQGLSSLQILDKVVKELSDRRMYVLLDHRSDCSAIS	143				
QY	125	ALWTSSEVSEATWISDQALQRYKGNPTVYVFDLHNEPDPACWGCSDSIDWELAAER	184				
Db	144	ELWYTDSEGEQWIDDLRFVAHRYTNVPGVIGLDYKNEPHGRATWGTGDPKTDMTVAEH	203				
QY	195	AGNAVLSTVNPILLFVEGQVSQNGDS-----YWGNGNLQAGQYFWLNVLP--NRLVY	235				
Db	204	AAAAILAEAPKWLIGVEGI-----GENPICSTIGHFWGENLEPMDCPT--LXVPANHLLL	257				
QY	236	SAHDYATSVYPTWFSDPTFNNMFGIWNKXWGLFNQNIAPVWLGEFTTL--QSTTDQ	293				
Db	258	MEHVYGPDPVYQPYFNSPDFENNMAAIWDQHFGRFPAKAGYA-NAIGFEGKYGEGDPRDI	316				
QY	294	TWLKTLVOYLPRPTAQYG-ADSFQWTFWSWNPDSGDTGGLXDDQQTV--DIVK	343				
Db	317	AWQNAFVDYL---ISIGVTDFAF---YMAANQSGDGTGCMWGNDDMTTPRDXKYK	363				

RESULT 15  
US-10-369-493-9262  
; Sequence 9262, Application US/10369493  
; Publication No. US2003023675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

Db 167 YTEFSEEDFTWIEVAKRFGKYNVIGADLKNEPHSVTSPAAAYTDGTGATWGMNPA 226  
Qy 176 IDWLAERAGNAVLNPNLLIFVEGVQSYN-----GDSYWGNNLQAGQYFV 226  
Db 227 TDWLAERIGKAILKVAPHWLI FVEGTQFTNPXTDSSYKMGYNWGMNMAVKDYPV- 285  
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFS DPT-PPNNMFGIWNKNGWYLFNQNIAPVWLGEFG 284  
Db 286 -NLPKNLVSPHYVGPVYNQYFPGAKGFPDNLFDIWHYHFGYVYKLELGYSVVIGFEG 344  
Qy 285 TTL---QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSNPDSDGTGGLKDDWQTV 339  
Db 345 GKYHGCDPRDVIWQNKLVDMW--IENKFCDF--YMSWNPDSGDTGGLQDDWTTI 397

## RESULT 10

US-10-369-493-1281  
; Sequence 1281, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 1281  
; LENGTH: 458  
; TYPE: PRT  
; ORGANISM: Pyrococcus horikoshii  
US-10-369-493-1281

Query Match 38.4%; Score 762; DB 15; Length 458;  
Best Local Similarity 45.0%; Pred. No. 3.9e-66;  
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPVRIAGINNFGEFTCNVYVHGLWSRDYRSMDOIKSLGYNTIRLPYSD 67  
Db 54 TSGBE-----TPHILFGVNWFGFETPNHVHGLWKRWEDMLLQIKSLGFNAIRLPFCT 107  
Qy 68 DILKPGTMENSINFRQMNQDLOGLTSQVMDKIVAYAGQIGLRIILDRHRPDCSGQSALW 127  
Db 108 ESVKPGTQPIGIDYSK-NPDLRGLDLSLOIMEKIIKKAGDLGIFVLLDYHRIGCTHIEPLW 166  
Qy 128 YTSSVSEATWISDLOALAQRYKGNTPVVGFDLHNEPHD-----PACWGCDDPS 175  
Db 167 YTEFSEEDFTWIEVAKRFGKYNVIGADLKNEPHSVTSPAAAYTDGTGATWGMNPA 226  
Qy 176 IDWLAERAGNAVLNPNLLIFVEGVQSYN-----GDSYWGNNLQAGQYFV 226  
Db 227 TDWLAERIGKAILKVAPHWLI FVEGTQFTNPXTDSSYKMGYNWGMNMAVKDYPV- 285  
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFS DPT-PPNNMFGIWNKNGWYLFNQNIAPVWLGEFG 284  
Db 286 -NLPKNLVSPHYVGPVYNQYFPGAKGFPDNLFDIWHYHFGYVYKLELGYSVVIGFEG 344  
Qy 285 TTL---QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSNPDSDGTGGLKDDWQTV 339  
Db 345 GKYHGCDPRDVIWQNKLVDMW--IENKFCDF--YMSWNPDSGDTGGLQDDWTTI 397

## RESULT 11

US-10-369-493-20347  
; Sequence 20347, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 20347  
; LENGTH: 458  
; TYPE: PRT  
; ORGANISM: Pyrococcus horikoshii  
US-10-369-493-20347

Query Match 38.4%; Score 762; DB 15; Length 458;  
Best Local Similarity 45.0%; Pred. No. 3.9e-66;  
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPVRIAGINNFGEFTCNVYVHGLWSRDYRSMDOIKSLGYNTIRLPYSD 67  
Db 54 TSGBE-----TPHILFGVNWFGFETPNHVHGLWKRWEDMLLQIKSLGFNAIRLPFCT 107  
Qy 68 DILKPGTMENSINFRQMNQDLOGLTSQVMDKIVAYAGQIGLRIILDRHRPDCSGQSALW 127  
Db 108 ESVKPGTQPIGIDYSK-NPDLRGLDLSLOIMEKIIKKAGDLGIFVLLDYHRIGCTHIEPLW 166  
Qy 128 YTSSVSEATWISDLOALAQRYKGNTPVVGFDLHNEPHD-----PACWGCDDPS 175  
Db 167 YTEFSEEDFTWIEVAKRFGKYNVIGADLKNEPHSVTSPAAAYTDGTGATWGMNPA 226  
Qy 176 IDWLAERAGNAVLNPNLLIFVEGVQSYN-----GDSYWGNNLQAGQYFV 226  
Db 227 TDWLAERIGKAILKVAPHWLI FVEGTQFTNPXTDSSYKMGYNWGMNMAVKDYPV- 285  
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFS DPT-PPNNMFGIWNKNGWYLFNQNIAPVWLGEFG 284  
Db 286 -NLPKNLVSPHYVGPVYNQYFPGAKGFPDNLFDIWHYHFGYVYKLELGYSVVIGFEG 344  
Qy 285 TTL---QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSNPDSDGTGGLKDDWQTV 339  
Db 345 GKYHGCDPRDVIWQNKLVDMW--IENKFCDF--YMSWNPDSGDTGGLQDDWTTI 397

## RESULT 12

US-10-369-493-17534  
; Sequence 17534, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 17534  
; LENGTH: 592  
; TYPE: PRT  
; ORGANISM: Xylella fastidiosa  
US-10-369-493-17534

Query Match 35.0%; Score 695.5; DB 15; Length 592;







QY 181 AERAGNAVLSVNPILLIFVEGVQSYNGDSYVWGGNLOAGAGQYPPVNLVNPRLVYSAHDY 240  
DB 222 AERAGNAVLSVNPILLIFVEGVQSYNGDSYVWGGNLOAGAGQYPPVNLVNPRLVYSAHDY 281  
QY 241 ATSVYPTQWFSPTFPNNPGIWNKNGYLFNQNIAPVWLGEFGTTLQSTTDTQTLV 300  
DB 282 ATSVYPTQWFSPTFPNNPGIWNKNGYLFNQNIAPVWLGEFGTTLQSTTDTQTLV 341  
QY 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQVTDVTKDGYLAPIKSSIFDPV 358  
DB 342 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQVTDVTKDGYLAPIKSSIFDPV 399

## RESULT 2

US-10-360-101-212  
; Sequence 212, Application US/10360101  
; Publication No. US20040009550A1  
; GENERAL INFORMATION:  
; APPLICANT: Moll, Gert N.  
; APPLICANT: Leenhouts, Cornelis J.  
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way  
; FILE REFERENCE: 2183-5673  
; CURRENT APPLICATION NUMBER: US/10/360,101  
; CURRENT FILING DATE: 2003-02-07  
; PRIOR APPLICATION NUMBER: EP 02077060.8  
; PRIOR FILING DATE: 2002-05-24  
; NUMBER OF SEQ ID NOS: 309  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 212  
; LENGTH: 397  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: sequence of cellulase  
US-10-360-101-212

Query Match 52.8%; Score 1048; DB 15; Length 397;  
Best Local Similarity 54.2%; Pred. No. 2e-94;  
Matches 193; Conservative 45; Mismatches 106; Indels 12; Gaps 4;

QY 4 GYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDSYRMLDQIKSLGYNITRL 63  
DB 37 GYHTQGNKIVDSGKEAFAFNGLNWFGLETPNTYTHGLMSRDMDDMLDQVKKGYNLRL 96  
QY 64 PYSDDIKPGTTPNSINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSQ 123  
DB 97 PYSNQLFDSRSRPSIDYHK-NPDLVGLNPQIMDKLIEKAGQGIQILDRHRPDCSQ 155  
QY 124 SALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183  
DB 156 SELWYTSQYPSRWISDWKVLADRYKKNPTVIGADLHNEPHGQASNGTGNASTDWRLAAQ 215  
QY 184 RAGNAVLSVNPILLIFVEGVQ---SYNGDSYVWGGNLOAGAGQYPPVNLVNPRLVYSAHDY 240  
DB 216 RAGNAVLSVNPILLIFVEGVQDHVNVGNSQYVWGGNLTGVANYPVVLVDVNPVYSPHDY 275  
QY 241 ATSVYPTQWFSPTFPNNPGIWNKNGYLFNQNIAPVWLGEFG---TTLQSTTDTQWLKT 298  
DB 276 PGVSSQWPNDFAPPNLPAIMDQWGYISKNIAPVLVGEFGCRNVDSLCPGKQWA 335  
QY 299 LVQYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQVTDVTKDGYLAPIKSS 354  
DB 336 LVHYI-----GANNLYFTYVSLNPNSGDTGGILLDDWTTWNPQDMLGRIMKPV 385

## RESULT 3

US-10-369-493-16379  
; Sequence 16379, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 16379  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-16379

Query Match 40.5%; Score 803.5; DB 15; Length 493;  
Best Local Similarity 46.9%; Pred. No. 3.5e-70;  
Matches 161; Conservative 61; Mismatches 102; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDSYRMLDQIKSLGYNITRLP 64  
DB 26 YSINNSRQIVDDSGKVQLKGVNVFETGTVHVMHGLWARNKDMIVQMGLGFNAVRLP 85  
QY 65 YSDDIKPGTTPNSINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSQS 124  
DB 86 FCPATLRSDTMPASIDYSR-NADLQGLTSLQILDKVTAEFNARGMYVLLDHHPTDCAGIS 144  
QY 125 ALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAER 184  
DB 145 ELWYTSVTEAQLADLRFVANRYKNVPYVGLDLKNEPHGATWGTGNAATDNKKAER 204  
QY 185 AGNAVLSVNPILLIFVEGVQ-----SYNGDSYVWGGNLOAGAGQYPPVNLVNP-NELVYSAH 238  
DB 205 GSAAVLAVAPKWLIAVEGITDNPVCSTNG-GIFWGGNLOPLACTP--LNIPANLLILAPH 261  
QY 239 DYATSVPTQWFSPTFPNNPGIWNKNGYLFNQNIAPVWLGEFGTTL--QSTTDTQWL 296  
DB 262 VYGDVVFQSVFNSFNPNPWPALWERHFGQFAGTH--ALLLGEFGKYGEGDARDKTWQ 319  
QY 297 KTLVQYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTV 339  
DB 320 DALVKYLR---SKGIN--QGFYWSNPNSGDTGGILRDDMTSV 357

## RESULT 4

US-10-369-493-16123  
; Sequence 16123, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS WITH IMPROVED PROPERTIES  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 16123  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-16123

Query Match 40.1%; Score 796.5; DB 15; Length 483;  
Best Local Similarity 46.6%; Pred. No. 1.7e-69;  
Matches 160; Conservative 61; Mismatches 103; Indels 19; Gaps 9;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:34:32 ; Search time 41.3333 Seconds  
(without alignments)  
2716.899 Million cell updates/sec

Title: US-09-997-504A-14

Perfect score: 1986  
Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSIPTDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1979	99.6	562	9	US-09-981-900B-5
2	1048	52.8	397	15	US-10-360-101-212
3	803.5	40.5	493	15	US-10-369-493-16379
4	796.5	40.1	483	15	US-10-369-493-16123
5	796.5	40.1	501	15	US-10-369-493-15740
6	796.5	40.1	518	15	US-10-369-493-15373
7	796.5	40.1	535	10	US-09-927-827-50
8	770	38.8	553	9	US-09-888-224-2
9	762	38.4	458	14	US-10-293-3448-2
10	762	38.4	458	15	US-10-369-493-1291
11	762	38.4	458	15	US-10-369-493-20347
12	695.5	35.0	592	15	US-10-369-493-17534
13	688.5	34.7	514	15	US-10-369-493-21618
14	667.5	33.6	369	15	US-10-369-493-9473
15	662.5	33.4	421	15	US-10-369-493-9262

16	583	29.4	472	15	US-10-369-493-8139	Sequence 8139, Ap
17	452.5	22.8	362	15	US-10-369-493-3080	Sequence 3080, Ap
18	271.5	13.7	353	15	US-10-369-493-15733	Sequence 15733, A
19	271.5	13.7	353	15	US-10-369-493-16119	Sequence 16119, A
20	271.5	13.7	582	10	US-09-927-827-48	Sequence 48, Appl
21	247.5	12.5	533	12	US-10-424-599-176232	Sequence 176232
22	243	12.2	375	15	US-10-369-493-12537	Sequence 12537, A
23	235	11.8	397	15	US-10-369-493-3152	Sequence 3152, Ap
24	196.5	9.9	438	12	US-10-425-114-66753	Sequence 66753, A
25	194.5	9.8	314	15	US-10-369-493-15726	Sequence 15726, A
26	194.5	9.8	314	15	US-10-369-493-16112	Sequence 16112, A
27	194.5	9.8	337	15	US-10-369-493-15359	Sequence 15359, A
28	188	9.5	517	12	US-10-107-431-259	Sequence 259, App
29	187	9.4	644	16	US-10-437-963-139883	Sequence 139883
30	182.5	9.2	917	16	US-10-437-963-181484	Sequence 181484
31	179.5	9.0	346	10	US-09-917-378-8	Sequence 8, Appli
32	175.5	8.8	375	10	US-09-917-378-3	Sequence 3, Appli
33	175.5	8.8	375	10	US-09-917-378-6	Sequence 6, Appli
34	175.5	8.8	762	10	US-09-917-378-1	Sequence 1, Appli
35	168	8.5	356	10	US-09-917-378-7	Sequence 7, Appli
36	164.5	8.3	558	16	US-10-437-963-185049	Sequence 185049
37	158.5	8.0	518	10	US-09-769-734-56	Sequence 56, Appl
38	154.5	7.8	722	12	US-10-282-122A-51993	Sequence 51993, A
39	153.5	7.7	1449	16	US-10-437-963-176100	Sequence 176100
40	152.5	7.7	533	12	US-10-424-599-150988	Sequence 150988
41	151.5	7.6	356	15	US-10-369-493-17756	Sequence 17756, A
42	149.5	7.5	268	12	US-10-424-599-284543	Sequence 284543
43	149.5	7.5	499	12	US-10-424-599-249378	Sequence 249378
44	148	7.5	217	16	US-10-437-963-112718	Sequence 112718
45	147.5	7.4	315	15	US-10-369-493-9177	Sequence 9177, Ap

#### ALIGNMENTS

#### RESULT 1

```
US-09-981-900B-5
; Sequence 5, Application US/09981900B
; Patent No. US2002013878A1
; GENERAL INFORMATION:
; APPLICANT: Sticklen, Masomeh B
; APPLICANT: Magbool, Shahina B
; APPLICANT: Dale, Bruce E
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULOSE WHICH DEGRAD
; TITLE OF INVENTION: AND CELLULOSE TO FERMENTABLE SUGARS
; FILE REFERENCE: MSU 4.1-539
; CURRENT APPLICATION NUMBER: US/09/981,900B
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 60/242,408
; PRIOR FILING DATE: 2000-10-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 562
; TYPE: PRI
; ORGANISM: Acidothermus cellulolyticus
US-09-981-900B-5
```

Query Match 99.6%; Score 1979; DB 9; Length 562;  
Best Local Similarity 99.7%; Pred. No. 5.4e-186;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	AGGYWHTSGREILDANNVPVRIAGINWFGPBTCTNVVHGLWSRDYRMLQIKSLGNT 60
Db	42	AGGYWHTSGREILDANNVPVRIAGINWFGPBTCTNVVHGLWSRDYRMLQIKSLGNT 101
QY	61	IRLPYSDDLKPGTWPNSINFRQMDLQGLTSLQWMDKIVAYAGQIGRLILDRHRDC 120
Db	102	IRLPYSDDLKPGTWPNSINFRQMDLQGLTSLQWMDKIVAYAGQIGRLILDRHRDC 161
QY	121	SGQSALWTSSVSEATWISDQLAQRKYGNPTVVVGFOLHNEPHDPACWGGDPSIDWRL 180
Db	162	SGQSALWTSSVSEATWISDQLAQRKYGNPTVVVGFOLHNEPHDPACWGGDPSIDWRL 221

Db 140 YNSIPGGLLVNGSNQAAIKEKYQKVVQWQIATKFSNVNRELIFESMNEVFD---GNYGNP 196  
Qy 175 SIDWRLAERAGNAVLSVNPILL--IFVEGVQSY--NGDSYW---WGNLQ-GAGQYP 224  
Db 197 -----NAAAYANLNAYNQIFVDTVQRTGGNNNARWLLIFGMNTNIDYTVGNYG 244  
Qy 225 VWL-----NVPN---RLVYSAHDY-----ATSVYP---QTFWSDP 253  
Db 245 PALPTDHFRRSAIFSSQKRIMISAHYSPWDFPAGEENGNTQWGAATNPKSKSTWQED 304  
Qy 254 TFPNNMFGIWNK--NWGYLFNQNIAPVWLGEFGT---TLOSTTDQWMLKTLVQYLAPTA- 307  
Db 305 YLNAQPKSMYDKFVTQGY-----PVVIGFSGSIDKTAYDSTNNVYRQAYAKAVTATAK 357  
Qy 308 QYGADSFQWTFWSWNPDSGDTG 329  
Db 358 KYGAVPVYW-----DNHGNG 372

Search completed: August 2, 2004, 16:38:35  
Job time : 15.6667 secs

Query Match 7.2%; Score 143; DB 4; Length 490;  
Best Local Similarity 23.1%; Pred. No. 3.2e-05;  
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;  
Qy 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCTVYVHVG-LWSRDY-RSMLDQIKSLGY 58  
Db 32 ANSGF-YVSGTTLYDANGNPFVNRGIN-----HGHAWYKDAQATTATEGIANTGA 79  
Qy 59 NTRLPYS-----DDILKPGTMPNSINFRQWV-----QDLOGLTSLQVMDKIYAY 103  
Db 80 NTVRIIVLSDGGOWTKDDI---HTVRNLISLAEDNHLVAVPEVHDATGYDSIASLNRVDY 136  
Qy 104 AGQIGLRILDRHRPDCSGSALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNE- 162  
Db 137 -----WIEMRSALI-----GKEDTVIINIINNEW 159  
Qy 163 --PHDPACWCGDPSIDWRLAERAGNAVLSVNPILLIFVEGVQSYNGDSYWGNNLOGA 220  
Db 160 FGSWEGDAWADG-----YKQAIPLRNA--GLNHTLMV-----DAAGW----- 195  
Qy 221 GOYPVVLNVPNRLVYSAHDYATSVY---PQ--TWFSPTFP---NNMFGIWNKNGYLFN 272  
Db 196 GQFP-----QSIHDI GREVFNADPQRTMFSIHMYEYAGGNASOV-RTNIDRVLN 244  
Qy 273 QNIAPVWLGEFGTTLQSTTDQWMLKTLVQYLAPTAQYGADSFQWTFWSWNPDSGDTGGIL 332  
Db 245 QDLALV-IGFEGH--RHTNGDVDEATMSY---SEQRGVG---WLAWSWKNG- 288  
Qy 333 KDDWQTVDTVKD 344  
Db 289 -PEWEYLDL.SND 299

## RESULT 15

US-09-797-464A-4  
; Sequence 4, Application US/09797464A  
; Patent No. 6630340  
; GENERAL INFORMATION:  
; APPLICANT: Wilting, Reinhard  
; APPLICANT: Bjornvad, Mads Eskelund  
; APPLICANT: Kauppinen, Markus Sakari  
; APPLICANT: Schuelein, Martin  
; TITLE OF INVENTION: Family 5 Xyloglucanases  
; FILE REFERENCE: 6073.200-US  
; CURRENT APPLICATION NUMBER: US/09/797,464A  
; CURRENT FILING DATE: 2002-02-19  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 400  
; TYPE: PRT  
; ORGANISM: Paenibacillus sp.  
US-09-797-464A-4

Query Match 7.2%; Score 142; DB 4; Length 400;  
Best Local Similarity 22.5%; Pred. No. 3e-05;  
Matches 86; Conservative 53; Mismatches 131; Indels 112; Gaps 23;  
Qy 3 GGYVHTSGREILDANNVPVRIAGINWFGFETCTVYVHGLWSRDYRSMLDQIKSLGYNTIR 62  
Db 48 GAGWNLGNQLEATVNGVFSETA---W-----GNPW-----TPELIKVKVKAAGFKTIR 92  
Qy 63 LP--YSDILKPGTMPN-SINFRQWMDLOGLTSLQVMDKIYAVAGQIGLRILDRHRPD 119  
Db 93 IPVSYLNHI---GAPNYTTINAALNLR-----VQTVVDYAYNEGLYVYVNIHGDC 139  
Qy 120 CSGQSALWYTSVSEATWISD-----LQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDP 174

```

US-08-525-697-2
; Sequence 2, Application US/08525697
; Patent No. 5795764
; GENERAL INFORMATION:
; APPLICANT: Christgau, Stephan
; APPLICANT: Andersen, Lene N
; APPLICANT: Kauppinen, Sakari
; APPLICANT: Helst-Hansen, Hans P
; APPLICANT: Dalboege, Henrik
; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5795764o No. 5795764disk of No. 5795764th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/525,697
; FILING DATE: 21-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Harrington, James J.
; REGISTRATION NUMBER: 38,711
; REFERENCE/DOCKET NUMBER: 4004.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0486/93
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-525-697-2

Query Match      8.9%; Score 176.5; DB 1; Length 377;
Best Local Similarity 23.1%; Pred. No. 1.3e-08;
Matches 81; Conservative 55; Mismatches 123; Indels 91; Gaps 19;

QY 24 AGIN-WFGEFTCNVYHGLWSRDYRSMDOIKSLGNTYRLPYSDDLKPGCTMPSINF 81
DB 50 AGTNSYMGFLTNDDVLMVMSQAFASDLKILRWGFDVNTKPTD-----GTWYQLHA 104
QY 82 ROMQDLQGLTSLQVMDKIVAYAGQIGIRILD--RHRPDCSGQSAL-----WY 128
DB 105 NGTSTINTGADGLQRLDYVYTSAEKYGKLIINFVNTDYGGMAYVAYGAAQTDYF 164
QY 129 TSSVSEATWISDLQALQAKYKGNVTYVGFDLHNEPHDPACWGCDPSIDWRLAAERAGNA 188
DB 165 TNTAIAQYAKYIKAVVSRYSSSAIAFAWELANEPR---CQGC-DTSVLNWSID-TSKY 219
QY 189 VLSVNPENLLIPVEGVQSVNGSDYWMGNLQAGQYPVV---LNVNRLVYSADHYAT-S 243
DB 220 IKSLDSKHLVTI-----GDE-CFGLDVSDSGPYTYGEGLNFTKNLGISIDFGTLH 271
QY 244 VYPQTFSDPTFPNNMFGIWNKNGWYLFNQIA-----PVMIGEGTT-----LQSTT 291
DB 272 LYPDSDWGTSYD-----WGNW---ITAAHAACKAVGKPCLLIEYGVTSNHCIVESPW 320
QY 292 DOT-----WLKTLVQYLRPTAQYGADSFQWTF-WSNPNPDSGDT 328
DB 321 QQTAGNATGSGDLYW-----QYGT-----TFSWGQSPDNCNT 353

```

```

RESULT 13
US-09-339-159B-4
; Sequence 4, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Bacillus
US-09-339-159B-4

Query Match      7.2%; Score 143; DB 4; Length 476;
Best Local Similarity 23.1%; Pred. No. 3.1e-05;
Matches 86; Conservative 43; Mismatches 111; Indels 132; Gaps 24;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDY-RSMLDQIKSLGY 58
DB 1 ANSGF-YVSGTTLYDANGNPFVVRGIN-----HGHAWYKDAQATTALEGIANPGA 48
QY 59 NTIRLPYS-----DDILKPGCTMPSINFQMN-----ODLQGLTSLQVMDKIVAY 103
DB 49 NTVRIVLSDGGQWTKDDI---HTVRNLISLAEDNHLVAVPEVHDATGYDSDIASLNRAVDY 105
QY 104 AGQIGIRILDRHRPDCSGQSALWYTSVSEATWISDLQALQAKYKGNPTVVGFDLHNE- 162
DB 106 -----WEMRSALI-----GKEDIVIIINIANEW 128
QY 163 --PHDPACWGCDPSIDWRLAAERAGNAVLSPNPLLIFVEGVQSVNGSDYWMGNLQGA 220
DB 129 FGSWEGDAWADG---YKQATPLRNA--GLNHTLMV-----DAAGW----- 164
QY 221 GQYPVVLNVPNRLVYSADHYATSVY---PQ--TWESDPTFP---NNMPGIWNKNGWYLFN 272
DB 165 GQFP-----QSIHDYGREVFNADPORNMTFESIHMIEYAGGNASQV-RTNIDRVIN 213
QY 273 QNIAPVWLGEFGTTTQSTTDTQTLKTLVQYLRPTAQYGADSFQWTFWSNPNPDSGDTGGIL 332
DB 214 QDLALV-IGEFCH--RHTNGDVDEATIMSY---SEQRGVG---WLAWSWKNG----- 257
QY 333 KDDWQTVDTVKD 344
DB 258 -PEWEYLDLSND 268

RESULT 14
US-09-339-159B-2
; Sequence 2, Application US/09339159B
; Patent No. 6566114
; GENERAL INFORMATION:
; APPLICANT: Kauppinen, Markus
; APPLICANT: Schulein, Martin
; APPLICANT: Schnorr, Kirk
; APPLICANT: Andersen, Lene
; APPLICANT: Bjornvad, Mads
; TITLE OF INVENTION: No. 6566114el Mannanases
; FILE REFERENCE: 5440.204-US
; CURRENT APPLICATION NUMBER: US/09/339,159B
; CURRENT FILING DATE: 1999-06-24
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2

```

QY 338 TVDTVKDGYLAP 349  
Db 562 TWDEQKYNFLKP 573

RESULT 10  
US-09-136-574A-43  
; Sequence 43, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, P.O. Box 457  
; CITY: Spring House  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/136,574A  
; FILING DATE: 19-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/932,571  
; FILING DATE: September 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 43:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1426 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No 6294366e  
; SEQUENCE DESCRIPTION: SEQ ID NO: 43:  
US-09-136-574A-43

Query Match 31.4%; Score 623; DB 3; Length 1426;  
Best Local Similarity 35.5%; Pred. No. 1.6e-50;  
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY 10 GRILLDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTIRLPYSDDI 69  
Db 1022 GNKIVDKGKPEVLTGVNWFNTGTNTVPDGVWNSCNLSALAEIARGFNLLRVPISAEI 1081  
QY 70 L---KPGTMEN-SINFRQMNQDLQGLTSLQWMDKIVAYAGQIGIRIILDRH--RPDCSQ 123  
Db 1082 ILNWSKGIYKPNINY-YVNPTEGLTSLVDFVVKTEKVGKIMLDIHSAKTDAMGH 1140  
QY 124 -SALWITSSVSEATWISDQLAQRYKGNVTGVDFLHNEPHDP-----ACWGGCDPS 175  
Db 1141 IYPVWYTDITTPEDYKACEWITERYKNDTIIVAFDLKNEPHGKPNQDSVFAKWDNSTDI 1200

QY 176 IDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWGGNLOGAGQ 222  
Db 1201 NNWKYAAETCAKRLAKPNPNLLIVIGIEIAYPKDDVTWTSKSSDYSTWGGNLRGVKK 1260  
QY 223 YPVVL-NVPNRLVYSADYATSVYPTW-----FSDPTFPNNMFGIWNKNWGYLFNQNIAP 277  
Db 1261 YPINLQYQKVYVSPHYGLYVYQPFYVPGFTKOTLYND---CWRDNWYIMDNGIAP 1317  
QY 278 VWLGERGTTILQSTTDQTLWLTIVQYLRPTAQYGAQSFQWTFWSWNPDSGDTGGILKDDWQ 337  
Db 1318 LLIGEGGYLDGGDNEKWMYLRDYII-----ENHIHTFWCYNANSNGDTGLGVGYDFS 1371  
QY 338 TVDTVKDGYLAP 349  
Db 1372 TWDEQKYNFLKP 1383

RESULT 11  
US-08-276-213-2  
; Sequence 2, Application US/08276213  
; Patent No. 5536655  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Steven  
; APPLICANT: Laymon, Robert  
; APPLICANT: Himmel, Michael  
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: National Renewable Energy Laboratory  
; STREET: 1617 Cole Boulevard  
; CITY: Golden  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80401-3393  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,213  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: O'Connor, Edna  
; REGISTRATION NUMBER: 29,252  
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303)231-1000  
; TELEFAX: (303)231-1098  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 38 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-08-276-213-2

Query Match 10.3%; Score 205; DB 1; Length 38;  
Best Local Similarity 97.4%; Pred. No. 8.2e-13;  
Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYV 38  
Db 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYV 38

RESULT 12





Query Match 38.8%; Score 770; DB 3; Length 553;  
Best Local Similarity 45.2%; Pred. No. 3.3e-65;  
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;  
QY 13 ILDANNVPVRIAGINWFGFETCYVYVHGLWSRDYRSMLEQIKSLGYNTRIRLPYSDDLKP 72  
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLPFCQSVKP 63  
QY 73 GTWPNISINFRQNDQLOGLTSLOVMDKIVAYAGQIGLRIILDRHRPDCSQSALWYTSV 132  
DB 64 GTMPTAIDYAK-NPDLQGLSDVQIMEKIIKKAGDLGIFVLDDYHRIGCNFIEPLWYDSF 122  
QY 133 SEATWISDLQALAOYKGNFTVVGFDLHNEPHDP-----ACWCGDPSIDWRL 180  
DB 123 SEQDYINTWVEAQRGKYWNVIAGDLKNEPHSSSPAPAAAYTDGSGATWGMGNATDNL 182  
QY 181 AAEAGNAVLSVNPNNLLIFVEGVQ-----SYN-GDSYVWGGNLQAGQYFVNLNP- 230  
DB 183 AAEIRIGRAILEVAPQWVIFVEGTQFTTPEIDGRYKWNHAWGNLGMVGRKYPV--NLPR 240  
QY 231 NRVYSADYATSVYPTWFSDP--TFPNNMPCGINKWGYLNFQNIAPVWLGEFGTTL- 287  
DB 241 DKVYSPQVYGVSEYDQPYF-DGEGFPDNLPEIWHHFGYVKLDLGYFVVIAGEGKYG 299  
QY 288 --QSTTDQTLVQLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWTV 339  
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YWSNPNSGDTGGILKDDWTTI 348

RESULT 6  
US-09-430-669-2  
; Sequence 2, Application US/09430669  
; Patent No. 6329187  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/430,669  
; FILING DATE: 28-Oct-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,572  
; FILING DATE: 22-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Herron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS: <unknown>  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-430-669-2

Query Match 38.8%; Score 770; DB 4; Length 553;  
Best Local Similarity 45.2%; Pred. No. 3.3e-65;  
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;  
QY 13 ILDANNVPVRIAGINWFGFETCYVYVHGLWSRDYRSMLEQIKSLGYNTRIRLPYSDDLKP 72  
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLPFCQSVKP 63  
QY 73 GTWPNISINFRQNDQLOGLTSLOVMDKIVAYAGQIGLRIILDRHRPDCSQSALWYTSV 132  
DB 64 GTMPTAIDYAK-NPDLQGLSDVQIMEKIIKKAGDLGIFVLDDYHRIGCNFIEPLWYDSF 122  
QY 133 SEATWISDLQALAOYKGNFTVVGFDLHNEPHDP-----ACWCGDPSIDWRL 180  
DB 123 SEQDYINTWVEAQRGKYWNVIAGDLKNEPHSSSPAPAAAYTDGSGATWGMGNATDNL 182  
QY 181 AAEAGNAVLSVNPNNLLIFVEGVQ-----SYN-GDSYVWGGNLQAGQYFVNLNP- 230  
DB 183 AAEIRIGRAILEVAPQWVIFVEGTQFTTPEIDGRYKWNHAWGNLGMVGRKYPV--NLPR 240  
QY 231 NRVYSADYATSVYPTWFSDP--TFPNNMPCGINKWGYLNFQNIAPVWLGEFGTTL- 287  
DB 241 DKVYSPQVYGVSEYDQPYF-DGEGFPDNLPEIWHHFGYVKLDLGYFVVIAGEGKYG 299  
QY 288 --QSTTDQTLVQLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWTV 339  
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YWSNPNSGDTGGILKDDWTTI 348

RESULT 7  
US-09-869-197-1  
; Sequence 1, Application US/09869197  
; Patent No. 6566113  
; GENERAL INFORMATION:  
; APPLICANT: TAKAYAMA, Masanori  
; APPLICANT: UMEDA, Kahoko  
; APPLICANT: KOYAMA, No. 6566113uto  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: TAKAYAMA-6  
; CURRENT APPLICATION NUMBER: US/09/869,197  
; CURRENT FILING DATE: 2001-06-25  
; PRIOR APPLICATION NUMBER: PCT/JP99/07009  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: JP 366237/1998  
; PRIOR FILING DATE: 1998-12-24  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Pyrococcus horikoshii OT3  
US-09-869-197-1

Query Match 38.4%; Score 762; DB 4; Length 430;  
Best Local Similarity 45.0%; Pred. No. 1.3e-64;  
Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;  
QY 8 TSGREILDANNVPVRIAGINWFGFETCYVYVHGLWSRDYRSMLEQIKSLGYNTRIRLPYS 67  
DB 26 TSGEE-----TPHLFGVNWFGFETPNVYVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 79  
QY 68 DILKGTWPNISINFRQNDQLOGLTSLOVMDKIVAYAGQIGLRIILDRHRPDCSQSALW 127  
DB 80 ESKVEGTQPIGIDYSK-NPDLRGLDSLQIMEKIIKKAGDLGIFVLDDYHRIGCTHIEPLW 138  
QY 128 YTSVSEATWISDLQALAOYKGNFTVVGFDLHNEPHD-----PACWCGDPS 175  
DB 139 YTEDPSEEDFINTWIEAKRFKYNWVIGADLKNEPHSVTSPPAAYTDGTGATWGMGNA 198  
QY 176 IDWRLAAERAGNAVLSVNPNNLLIFVEGVQSYN-----GDSYVWGGNLQAGQYFV 226

Query Match 38.8%; Score 770; DB 1; Length 553;  
Best Local Similarity 45.2%; Pred. No. 3.3e-65;  
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;

QY 13 ILDANNVPIRIAGNWFETCNVYVHGLWSRDYRSLDQIKSLGYNTRILPYSDDLKP 72  
DB 4 VATGEETPIHLFGVNWFFETPNVYVHGLWSRWEDMLLQIKSLGYNTRILPYSDDLKP 63  
QY 73 GTPMNSINFRQMNQDLQGLTSLOWMDKIVAYAGQIGLRIILDRHRPCDQSGSALWYTSV 132  
DB 64 GTMPTAIDYAK-NPDQLGLDSVQIMEKIIKKAGDLGIFVLDDYHRIKGFIEPLWYDTSF 122  
QY 133 SEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----ACWGCDDPSIDWRL 180  
DB 123 SEQDIINTWVEVAQRFGKYNNVIGADLKNEPHSSSPAPAAAYTDCSGATWGMGNATDWNL 182  
QY 181 AARAGNAVLNVNPNLLIFVEGVQ-----SYN-GDSYWMGNLQAGQYPPVNLNVP- 230  
DB 183 AAEIRIGRAILEVAPQWVIFVEGTQFTTPEIDGRYKWHNAWGNLGMVRYKYPV--NLPR 240  
QY 231 NRLVYSADHYATSVYPTWFSDP--TPNNMNGIWNKNWGLFNQNIAPVWLGEFGTTL- 287  
DB 241 DKVYSPQVYGVSEYVDQYF-DPGEFPDNLPEIWHHFGYVLDLGYPPVIGFEGKYG 299  
QY 288 --QSTTDQTLKTLVQYLRPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTV 339  
DB 300 HGGDRDVTWQNKIIDNM--IQNKFCDFP---YWSNPNNSGDTGGILKDDWTTI 348

## RESULT 4

US-09-066-544-2  
; Sequence 2, Application US/09066544  
; Patent No. 6001984  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/066,544  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,572  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Herron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-09-066-544-2

Query Match 38.8%; Score 770; DB 3; Length 553;  
Best Local Similarity 45.2%; Pred. No. 3.3e-65;  
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;

QY 13 ILDANNVPIRIAGNWFETCNVYVHGLWSRDYRSLDQIKSLGYNTRILPYSDDLKP 72  
DB 4 VATGEETPIHLFGVNWFFETPNVYVHGLWSRWEDMLLQIKSLGYNTRILPYSDDLKP 63  
QY 73 GTPMNSINFRQMNQDLQGLTSLOWMDKIVAYAGQIGLRIILDRHRPCDQSGSALWYTSV 132  
DB 64 GTMPTAIDYAK-NPDQLGLDSVQIMEKIIKKAGDLGIFVLDDYHRIKGFIEPLWYDTSF 122  
QY 133 SEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----ACWGCDDPSIDWRL 180  
DB 123 SEQDIINTWVEVAQRFGKYNNVIGADLKNEPHSSSPAPAAAYTDCSGATWGMGNATDWNL 182  
QY 181 AARAGNAVLNVNPNLLIFVEGVQ-----SYN-GDSYWMGNLQAGQYPPVNLNVP- 230  
DB 183 AAEIRIGRAILEVAPQWVIFVEGTQFTTPEIDGRYKWHNAWGNLGMVRYKYPV--NLPR 240  
QY 231 NRLVYSADHYATSVYPTWFSDP--TPNNMNGIWNKNWGLFNQNIAPVWLGEFGTTL- 287  
DB 241 DKVYSPQVYGVSEYVDQYF-DPGEFPDNLPEIWHHFGYVLDLGYPPVIGFEGKYG 299  
QY 288 --QSTTDQTLKTLVQYLRPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTV 339  
DB 300 HGGDRDVTWQNKIIDNM--IQNKFCDFP---YWSNPNNSGDTGGILKDDWTTI 348

## RESULT 5

US-08-951-086-2  
; Sequence 2, Application US/08951086  
; Patent No. 6074867  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/951,086  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/651,572  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Herron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-951-086-2

Query Match 99.6%; Score 1979; DB 1; Length 358;  
Best Local Similarity 99.7%; Pred. No. 2.5e-181;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNT 60  
DB 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNT 60

QY 61 IRLPYSDDILKPGTWPNSINFRQNDLOGLTSLQVMDKIVAYAGQIGLRILIDHRPDC 120  
DB 61 IRLPYSDDILKPGTWPNSINFRQNDLOGLTSLQVMDKIVAYAGQIGLRILIDHRPDC 120

QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180

QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYWWGNLQAGQYPVVLNPNRLVYSAHDY 240  
DB 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYWWGNLQAGQYPVVLNPNRLVYSAHDY 240

QY 241 ATSVPQTWFSPTFPNNMFGIWNKNGYLFNQNTAPVWLGEFGTTLOSTTDQTLWKTIV 300  
DB 241 ATSVPQTWFSPTFPNNMFGIWNKNGYLFNQNTAPVWLGEFGTTLOSTTDQTLWKTIV 300

QY 301 QYLRPTAQYGADSFQWTFWSPNPDSDGTGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
DB 301 QYLRPTAQYGADSFQWTFWSPNPDSDGTGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

## RESULT 2

US-08-276-213-3  
; Sequence 3, Application US/08276213  
; Patent No. 5536655  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Steven  
; APPLICANT: Layton, Robert  
; APPLICANT: Himmel, Michael  
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: National Renewable Energy Laboratory  
; STREET: 1617 Cole Boulevard  
; CITY: Golden  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80401-3393  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,213  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: O'Connor, Edna  
; REGISTRATION NUMBER: 29,252  
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08  
; TELEPHONE: (303)231-1000  
; TELEFAX: (303)231-1098  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 521 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; US-08-276-213-3

Query Match 99.6%; Score 1979; DB 1; Length 521;  
Best Local Similarity 99.7%; Pred. No. 4.4e-181;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNT 60  
DB 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNT 60

QY 61 IRLPYSDDILKPGTWPNSINFRQNDLOGLTSLQVMDKIVAYAGQIGLRILIDHRPDC 120  
DB 61 IRLPYSDDILKPGTWPNSINFRQNDLOGLTSLQVMDKIVAYAGQIGLRILIDHRPDC 120

QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180

QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYWWGNLQAGQYPVVLNPNRLVYSAHDY 240  
DB 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYWWGNLQAGQYPVVLNPNRLVYSAHDY 240

QY 241 ATSVPQTWFSPTFPNNMFGIWNKNGYLFNQNTAPVWLGEFGTTLOSTTDQTLWKTIV 300  
DB 241 ATSVPQTWFSPTFPNNMFGIWNKNGYLFNQNTAPVWLGEFGTTLOSTTDQTLWKTIV 300

QY 301 QYLRPTAQYGADSFQWTFWSPNPDSDGTGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
DB 301 QYLRPTAQYGADSFQWTFWSPNPDSDGTGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

## RESULT 3

US-08-651-572-2  
; Sequence 2, Application US/08651572  
; Patent No. 5789228  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,572  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Heiron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; US-08-651-572-2

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: August 2, 2004, 16:31:56 ; Search time 14.6667 Seconds  
(without alignments)  
1260.143 Million cell updates/sec

Title: US-09-997-504A-14  
Sequence: 1 AGGGYHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDFV 358  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AX:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES									
Result No.	Score	Query		Length	DB	ID	Description		
		Match	%						
1	1979	99.6	358	1	US-08-604-913B-11		Sequence 11, Appl		
2	1979	99.6	521	1	US-08-276-213-3		Sequence 3, Appl		
3	770	38.8	553	1	US-08-651-572-2		Sequence 2, Appl		
4	770	38.8	553	3	US-09-066-544-2		Sequence 2, Appl		
5	770	38.8	553	3	US-08-951-086-2		Sequence 2, Appl		
6	770	38.8	553	4	US-09-430-669-2		Sequence 2, Appl		
7	762	38.4	430	4	US-09-869-197-1		Sequence 1, Appl		
8	762	38.4	458	4	US-09-869-197-5		Sequence 5, Appl		
9	623	31.4	616	3	US-09-136-574A-47		Sequence 47, Appl		
10	623	31.4	1426	3	US-09-136-574A-43		Sequence 43, Appl		
11	205	10.3	38	1	US-08-276-213-2		Sequence 2, Appl		
12	176.5	8.9	377	1	US-08-525-697-2		Sequence 2, Appl		
13	143	7.2	476	4	US-09-339-159B-4		Sequence 4, Appl		
14	143	7.2	490	4	US-09-339-159B-2		Sequence 2, Appl		
15	142	7.2	400	4	US-09-797-464A-4		Sequence 4, Appl		
16	138	6.9	348	4	US-09-797-464A-7		Sequence 7, Appl		
17	138	6.9	468	4	US-09-485-648-4		Sequence 4, Appl		
18	138	6.9	468	4	US-09-503-565-4		Sequence 4, Appl		
19	138	6.9	468	4	US-09-485-649-4		Sequence 4, Appl		
20	138	6.9	468	4	US-09-339-159B-8		Sequence 8, Appl		
21	138	6.9	493	4	US-09-485-648-2		Sequence 2, Appl		
22	138	6.9	493	4	US-09-503-565-2		Sequence 2, Appl		
23	138	6.9	493	4	US-09-485-649-2		Sequence 2, Appl		
24	138	6.9	493	4	US-09-339-159B-6		Sequence 6, Appl		
25	137	6.9	395	4	US-09-797-464A-2		Sequence 2, Appl		
26	135	6.8	363	4	US-09-797-464A-11		Sequence 11, Appl		
27	132.5	6.7	663	4	US-09-134-078-61		Sequence 61, Appl		

28	132.5	6.7	680	4	US-09-134-078-25	Sequence 25, Appl
29	131	6.6	24	1	US-08-276-213-1	Sequence 1, Appl
30	129	6.5	331	4	US-09-339-159B-12	Sequence 12, Appl
31	126	6.3	320	4	US-09-339-159B-22	Sequence 22, Appl
32	120	6.0	331	3	US-08-849-751-4	Sequence 4, Appl
33	120	6.0	331	4	US-08-478-816-4	Sequence 4, Appl
34	114.5	5.8	429	1	US-08-745-977-4	Sequence 4, Appl
35	114.5	5.8	429	3	US-09-040-699A-4	Sequence 4, Appl
36	112.5	5.7	360	4	US-09-134-078-27	Sequence 27, Appl
37	110.5	5.6	490	1	US-08-672-571A-1	Sequence 1, Appl
38	110.5	5.6	490	1	US-08-672-571A-1	Sequence 1, Appl
39	109	5.5	317	2	US-09-066-075-2	Sequence 2, Appl
40	109	5.5	317	2	US-08-518-615A-2	Sequence 2, Appl
41	109	5.5	317	3	US-09-951-889-2	Sequence 2, Appl
42	109	5.5	317	3	US-09-472-857-2	Sequence 2, Appl
43	108	5.4	327	2	US-08-169-948B-16	Sequence 16, Appl
44	108	5.4	327	2	US-08-448-873-16	Sequence 16, Appl
45	108	5.4	327	3	US-08-382-452D-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1  
US-08-604-913B-11  
; Sequence 11, Application US/08604913B  
; Patent No. 5712142  
; GENERAL INFORMATION:  
; APPLICANT: Acney, William S.  
; APPLICANT: Thomas, Steven R.  
; APPLICANT: Himmel, Michael E.  
; APPLICANT: Baker, John O.  
; APPLICANT: Chou, Yat-Chen  
; TITLE OF INVENTION: METHOD FOR INCREASING  
; TITLE OF INVENTION: THERMOSTABILITY IN CELLULOSE ENZYMES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: National Renewable Energy Laboratory  
; STREET: 1617 Cole Boulevard  
; CITY: Golden  
; STATE: CO  
; COUNTRY: U.S.A.  
; ZIP: 80401-3393  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBC PC compatible  
; SOFTWARE: ASC II (DOS) text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/604,913B  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/276,213  
; FILING DATE: 15-070-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Edna M. O'Connor  
; REGISTRATION NUMBER: 29,252  
; REFERENCE/DOCKET NUMBER: 95-56  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/384-7573  
; TELEFAX: 303/384-7499  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 358 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; FEATURE:  
; NAME/KEY: E1-CAT  
; US-08-604-913B-11



PT Polypeptides with heat-resistant cellobiohydrolase activity for efficient  
 PT breakdown of cellulose biomass.

XX Disclosure; Page 45-47; 50pp; Japanese.

XX This invention describes a novel polypeptide originating in *Pyrococcus*  
 CC horikoshii OT3 which has cellobiohydrolase activity. The polypeptide of  
 CC the invention is capable of decomposing poly(D-glucopyranose) having beta  
 CC -1,4 bonds and can be used for the efficient and straightforward  
 CC breakdown of cellulose biomass to glucose. This sequence represents a P.  
 CC horikoshii OP3 cellobiohydrolase associated protein described in the  
 CC method of the invention

XX Sequence 458 AA;

Query Match 38.4%; Score 762; DB 3; Length 458;  
 Best Local Similarity 45.0%; Pred. No. 4e-63;  
 Matches 161; Conservative 48; Mismatches 109; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRSMLDQIKSLQYNTIRLPYS 67  
 Db 54 TSGEE-----TPHLFGVNWFGFETPNVHVGWKNWEDMLLQIKSLGFNAIRLPFCT 107  
 Qy 68 DILKPGTWPNSINFRQMNQDLOGITSLQVMDKIVAVAGQIGLRIILDRHRPDCSGOSALW 127  
 Db 108 ESKVPGTQPIDYDK-NPDLRLGLDSLQIMEKIKKAGDLGIFVLLDYHRIGCTHIEPLW 166  
 Qy 128 YTSVSSATWISDLQALQARYKGNPTVVGFDLNEPHD-----PACWGGCDPS 175  
 Db 167 YTEDFSEDFINTWIEVAKRFGKYNVIGADLKNEPHSVTSPPAAVTDGTGTWGMGNEA 226  
 Qy 176 IDNELAERAGNAVLSVNPMLLIFVEGVQSYN-----GDSYMWGKNLQAGQYPPV 226  
 Db 227 TDNLAERIGKALKVAPHLIFVEGTQFTNPKTSSYKGINAMWGNLMAVKDYPV- 285  
 Qy 227 LNVP-NRLVYSAHDYATSVYPTQWFSPT-FPNMNGINNNKNGYLFNQNIAPVWLGEGF 284  
 Db 286 -NLPNKLAVSPHYGPDVYNQYFGPAKGFDPNLPDINHYHFGYVKLELGSVWIGEGF 344  
 Qy 285 TTL---QSTDTQWLKTLVQRLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWTV 339  
 Db 345 GYGHGDPDRDVIWQNLVDWM--IENKFCDFP---YWSNPDSDGTGGILQDDWTI 397

RESULT 14  
 AAB96264  
 ID AAB96264 standard; protein; 514 AA.

XX AAB96264;

XX 29-OCT-2001 (first entry)

XX Putative P. abyssi endoglucanase #2.

XX Hyperthermophilic archaeon; hyperthermophilic protein.

XX *Pyrococcus abyssi*.

XX FR2792651-A1.

XX 27-OCT-2000.

XX 21-APR-1999; 99FR-00005034.

XX 21-APR-1999; 99FR-00005034.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX (IFRE-) IFREMER INST FR RECH EXPL MER.

XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX Querehrou J, Weissenbach J, Saurin W, Heilig R;

XX WPI; 2001-126236/14.

XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins  
 PT useful in industry.

XX Claim 7; Page 921-922; 1657pp; French.

XX The present invention relates to the genomic sequence of *Pyrococcus*  
 CC abyssi (see AAF6431 and AAH41223-7) and P. abyssi proteins. P. abyssi is  
 CC a hyperthermophilic archaeon, which is isolated from deep-sea  
 CC hydrothermal vents. The present invention is one such P. abyssi protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade. Note: This patent is in the same patent family as  
 CC WO2000065062, which contains additional sequences as shown in AAB99132-  
 CC AAB99143, AAH75903-AAH75920 and AAG66436

XX Sequence 514 AA;

Query Match 34.7%; Score 688.5; DB 4; Length 514;  
 Best Local Similarity 41.6%; Pred. No. 4.3e-56;  
 Matches 148; Conservative 58; Mismatches 121; Indels 29; Gaps 10;

Qy 7 HTSGREILDANNV-----PVRIAGINWFGFETCNVYVHGLWSRDRSMLDQIKSLGNT 60  
 Db 30 YTAENGIIIFVQNVTTGEEKKPLYLHGVSWFGFELKDHVYGLDKRNWKDILKDVRLGFNA 89  
 Qy 61 IRLPYSDIILKPGTWPN--SINFRQMNQDLOGITSLQVMDKIVAVAGQIGLRIILDRHRP 118  
 Db 90 IRLPFCSSIRPDTPSPERINY-ELNPDALKLTSLEIMEKIIIVANSIGLYLLDVHRI 148  
 Qy 119 DCSGOSALWYSSVSSEATWISDLQALQARYKGNPTVVGFDLNEPHDPACWGGCDPSIDW 178  
 Db 149 GCEEIPLWYNTENYSEEQYIKDWIFLAKRFGKYPNVIGADIKNEPHGAGWGTGDER-DF 207  
 Qy 179 RLAAERAGNAVLSVNPMLLIFVEGVQ-----SYNGDSYV--WGNLQAGQYPPVL 227  
 Db 208 RLFAEKVGRILKVAAPHLIFVEGTQYTHVNPINDEIIEKKGWTFWGNLWGVKDYVRL 267  
 Qy 228 NVENRLVYSAHDYATSVYPTQWFSPT-FPNMNGINNNKNGYLFNQNIAPVWLGEGTTL 287  
 Db 268 -PRGVVYSPHYGPDVYNQYFGPAKGFDPNLPDINHYHFGYVKLELGSVWIGEGN 325  
 Qy 286 QSTDTQWLKTLVQRLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWTV 343  
 Db 326 EG-LDKVQDAPVFWLKKLY-----NFFWCLNPESDGTGGIFLDDKVTNWEK 375

RESULT 15  
 AAY13494

ID AAY13494 standard; protein; 616 AA.

XX AAY13494;

XX 30-JUL-1999 (first entry)

XX Truncated cellulase Cel E3/B5.

XX Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

XX Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;

XX cotton-containing fabric; stonewashing.

XX Unidentified.

XX EP921188-A2.

XX 09-JUN-1999.

XX 15-SEP-1998; 98EP-00810919.

XX 19-SEP-1997; 97US-00932571.

XX (CLN) CLARIANT FINANCE BVI LTD.

the invention is a thermostable enzyme which is heat stable, catalyses the enzymatic hydrolysis of cellulose and is able to renature and regain activity after exposure to temperatures of 60-105 degrees Centigrade. The polynucleotide and polypeptide can be used to raise an antibody, generate a variant and as a probe for isolating or identifying other endoglucanase genes. Also disclosed is a method for comparing a first sequence to a reference sequence, and for identifying a feature, through the use of a computer program. The endoglucanase has increased activity and stability at increased pH and temperature. The sequence presented is the *T. maritima* endoglucanase protein

```

XX
SQ Sequence 553 AA;
Query Match 38.8%; Score 770; DB 5; Length 553;
Best Local Similarity 45.2%; Pred. No. 9.1e-64;
Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;

QY 13 ILDANNVPRIAGINWGFETCTNVVHGLWSRDYRSLMDQIKSIGYNTIRLPYSDDLKP 72
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 VATGEETPIHLFGYNWGFETPNVVHGLWSRNWEDMLQIKSLGFNAIRLPFCQTQVKP 63
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 73 GTVPNSINFRQMNDLOGLTSLQYMDKIVAVAGOIGLRRIILDRHRPCSGQSALWYTSV 132
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 GTVPTALDYAK-NPDLQGLDSVQIMKLIKAGDLGIFVLJDYHRIICGNFTPEIPLYTDSF 122
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 133 SEATWISDLQALQRYKGNPTVVGFDLHNEHPD-----ACWGCGDSPIDWRL 180
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 SEQDYINTWVEVAQRFQKYNVIGADLKNEPHSSSPAPAYTDGSGATGWMGNATDWNL 182
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 AERAGHANVLNVNPLLIIFVEGVO-----SYN-GDSYWGNGNLOAGQGVVVLNVP- 230
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 AAEIRIGRAILEVAPQWIFVEGTOFTTPEIDGRYKWNAGWNGNLMGVRKYPV--NLPR 240
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 NRLVYSADHVATSVYPTWFSDP--TPPNNNPGIWNKNWGLFNQNTAPVWLGEFGITL- 287
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 DKVVISQVGVSEVYDQYF-DPREGFDNLPETLYWHHFGYVVKLLDGYPVVIGFEGKYG 299
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 --QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQTV 339
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 300 HGGDPRPTVWONKIIDWV--TONKFCDFP---YWSWNPNSGDTGGILKDDWTTI 348
: : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 12	
AAAB10344	
ID	AAAB10344 standard; protein; 430 AA.
XX	
XX	AAAB10344;
XX	
XX	
DT	24-NOV-2000 (first entry)
XX	
DE	P. Horikoshii OT3 cellobiohydrolase associated protein.
XX	
KW	Cellobiohydrolase; poly(D-glucopyranose) decomposition; glucose;
KW	cellulose breakdown.
XX	
OS	Pyrococcus horikoshii.
XX	
XX	WO200039288-A1.
XX	
PD	06-JUL-2000.
XX	
XX	
PF	14-DEC-1999; 99WO-JP007009.
XX	
PR	24-DEC-1998; 98JP-00366237.
XX	
PA	(TAKI ) TAKARA SHUZO CO LTD.
XX	
PI	Takayama M, Umeda K, Koyama N, Asada K, Kato I;
XX	
DR	WPI; 2000-452391/39.
DR	N-PSDB; AAA71327.
XX	
PT	Polypeptides with heat-resistant cellobiohydrolase activity for efficient

breakdown of cellulose biomass.  
Claim 1; Page 41-43; 50pp; Japanese.  
This invention describes a novel polypeptide originating in *Pyrococcus horikoshii* OP3 which has cellobiohydrolase activity. The polypeptide of the invention is capable of decomposing poly(D-glucopyranose) having beta -1,4 bonds and can be used for the efficient and straightforward breakdown of cellulose biomass to glucose. This sequence represents the P. horikoshii OP3 cellobiohydrolase associated protein described in the method of the invention  
Sequence 430 AA;  
SQ

Query Match.	38.4%;	Score	762;	DB 3;	Length	430;			
Best Local Similarity	45.0%;	Pred.	No. 3.6e-63;						
Matches	161;	Conservative	48;	Mismatches	109;	Indels	40;	Gaps	10;
Qy	8	TSGREILDANNVPIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTRILRPSD	67						
Db	26	TSGE-----TTHUFGVNWFGFETPNVYVHGLWKNWEDMLQIKSLGFNAIRLFPCT	79						
Qy	68	DIUKPGTMPNSINFRQMNQLOGJTSLQVNDKIVAYAGQIGLRIILDRHRPDCSGQSALW	127						
Db	80	ESVKPGTQPIGIDYSK-NPDLRGLDSLQIWEKIIKKAAGDLGIFVLLDVYHRIGCTHIEPLW	138						
Qy	128	YTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWGCDGPS	175						
Db	139	YTEDFSEDFINTWIEVAKFEGKYNNVIGADLKNEPHSVTSPAAAYTDGTATWGNCPA	198						
Qy	176	IDWRLAERAGNAVLSYNPNLLIFVEGVQSYN-----GDSYMWGNLQAGAQPVV	226						
Db	199	TDNMLAERIGKAILKVAPHLIFVEGTQTNPKTDSYKWKYNAMWGNLMAVKDIPV-	257						
Qy	227	LNVP-NELVYSAHDYATSVYPQWFSDDPT-FPNNMPCIMWNKNGYLFNQNIAPVWLGEFG	284						
Db	258	NLPNRNKLVSYPHYGPDVYNQYFPGAQFPDNLPDYVHHFGYVKLELGSYSVVIIEFG	316						
Qy	285	TTL---QSTTDQWLKTLVOYLAPRTAGYGADSPQWTFWSNPNDSGDTGGILKDDQWTV	339						
Db	317	GKYGHGDDPRDVTWONKLVDMW---IENKCFDFF---YWSWNPDSGDTGGILQDDWTTI	369						

RESULT l3	
AABl0345	
ID	AABl0345 standard; protein; 458 AA.
XX	
AC	AABl0345;
XX	
DT	24-NOV-2000 (first entry)
XX	
DE	P. horikoshii Otr3 cellobiohydrolase associated protein #2.
XX	
KW	Cellobiohydrolase; poly(D-glucopyranose) decomposition; glucose;
KW	cellulose breakdown.
XX	
OS	Pyrococcus horikoshii.
XX	
PN	WC200039288-Al.
XX	
PD	06-JUL-2000.
XX	
PF	14-DEC-1999; 99WO-JP007009.
XX	
PR	24-DEC-1998; 98JP-00366237.
XX	
PA	(TAKI ) TAKARA SHUZO CO LTD.
XX	
PI	Takayama M, Umeda K, Koyama N, Asada K, Kato I;
XX	
DR	WPI; 2000-452391/39.
DR	N-ESDB; AAA71330.
XX	

polypeptides with heat-resistant cellobiohydrolase activity for efficient

Db 183 AAEIRGRAILEVAPQWVIFVEGTQFTPEIDGRYKGNHAWGNLGMVKYKYPV--NLPR 240  
 Qy 231 NRVLSAHDYATSYPTWFSDDP--TFPNNMPCIMKNGWYLFNQNIAPVWLGEFTTL- 287  
 Db 241 DKLVSPQVYGPVYDQYF--DPGEGFPDNLPEIWHHFGYVXLDLGPVWVGEFGKYG 299  
 Qy 288 --QSTDTQWLKTLVQYLRFTAQYGADSFQWTFWNNPDSGDTGGILKDDWQTV 339  
 Db 300 HGGDPRDVTQNKIIDMW--IQNKPCDF--YMSWNPNSGDTGGILKDDWTTI 348

RESULT 10  
 AAW34985  
 ID AAW34985 standard; protein; 841 AA.  
 XX AC AAW34985;  
 XX 27-AUG-2003 (revised)  
 DT 21-MAY-1998 (first entry)  
 XX Archaeobacterial thermostable endoglucanase.  
 DE Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;  
 KW beta-1,4-glycosidic bond; hydrolysis; saccharification;  
 KW archaeal bacterium; thermostable enzyme; thermophilic.  
 XX archaeon.  
 OS Unidentified.  
 XX WO9744361-A1.  
 XX 27-NOV-1997.  
 XX 22-MAY-1997; 97WO-US008793.  
 XX 22-MAY-1996; 96US-00651572.  
 XX (RECO-) RECOMBINANT BIOCATALYSIS INC.  
 XX Lam DE, Mathur EJ;  
 XX WPI; 1998-018435/02.  
 XX N-ESDB; AAT94193.  
 XX Endo:glucanase(s), preferably from archaeal bacterium, ABPII 1a - useful  
 PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic  
 PT bonds in cellulose.  
 XX Claim 1; Fig 1A; 164pp; English.  
 XX This protein comprises a thermostable 60.9 kDa endoglucanase of archaeal  
 CC bacterial strain ABPII1a, a thermophilic isolate of a marine hydrothermal  
 CC vent. The endoglucanase is capable of degrading carboxymethylcellulose  
 CC and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It can be  
 CC produced from native cells or from recombinant host cells, especially  
 CC prokaryotic host cells transformed with a plasmid or virus-derived vector  
 CC including the endoglucanase DNA (see AAT94193). 23 Other, homologous  
 CC endoglucanases (see AAW34986-W35008) are also claimed. The endoglucanase  
 CC can be used to degrade cellulose for the conversion of plant biomass into  
 CC fuels and chemicals, for use in detergents, textiles, animal feed, waste  
 CC treatment, and in the fruit juice and brewing industries for the  
 CC clarification and extraction of juices. (Updated on 27-AUG-2003 to  
 CC correct OS field.)  
 XX Sequence 841 AA;  
 XX Query Match 38.8%; Score 771; DR 2; Length 841;  
 XX Best Local Similarity 45.2%; Pred. No. 1.3e-63;  
 XX Matches 160; Conservative 52; Mismatches 106; Indels 36; Gaps 11;  
 Qy 13 ILDANNVPRIAGINWGFETCNVVGWLSRSDYRSMIDQIKSYNTIRLPYSDDLKP 72

Db 4 VATGEETPIHLFGVNWFGFETPNVYVHGLMSRNWEDMLQIKSLGFNAIRLPFCTQSVKP 63  
 Qy 73 GTMENSINFRQWQDLOGLTSLQWMDXIVAYAGOIGLRILILDRHRPCDSCGOSALWYTSV 132  
 Db 64 GTMPTAIDYAK-NPDLOGLDSVQIMEKLIKAGDLGIFVLLDYHHRIGCNFTEPLWYTDSP 122  
 Qy 133 SEATWISDLOALAQRYKGNPTWVGFDLHNEPHDP-----ACWCGDPSIDWRL 180  
 Db 123 SEQDINTWVEVAQRFGKYNNVIGADLKNEPHSSPPAPAAVTDGSGATWGMNNATDNL 182  
 Qy 181 AERAGNAVLSNENLLIFVEGVQ-----STN-GDSYMWGNLQAGQYFVVLNVP- 230  
 Db 183 AERIGRAILEVAPQWVIFVEGTQFTPEIDGRYKGNHAWGNLGMVKYKYPV--NLPR 240  
 Qy 231 NRVLSAHDYATSYPTWFSDDP--TFPNNMPCIMKNGWYLFNQNIAPVWLGEFTTL- 287  
 Db 241 DKLVSPQVYGPVYDQYF--DPGEGFPDNLPEIWHHFGYVXLDLGPVWVGEFGKYG 299  
 Qy 288 --QSTDTQWLKTLVQYLRFTAQYGADSFQWTFWNNPDSGDTGGILKDDWQTV 339  
 Db 300 HGGDPRDVTQNKIIDMW--IQNKPCDF--YMSWNPNSGDTGGILKDDWTTI 348

RESULT 11  
 ABG70759  
 ID ABG70759 standard; protein; 553 AA.  
 XX AC ABG70759;  
 XX 09-DEC-2002 (first entry)  
 DT T. maritima endoglucanase protein.  
 DE Endoglucanase; enzyme; cellulose; plant cell wall; homopolysaccharide;  
 KW D-glucose; beta configuration; beta 1-4 glycosidic bond; hydrolysis;  
 KW 1,4-beta-D-glucan glucanohydrolase; exoglucanase;  
 KW 1,4-beta-D-glucan cellobiohydrolase; thermostable; catalysis; antigen.  
 XX Thermotoga maritima.  
 XX US2002120118-A1.  
 XX 29-AUG-2002.  
 XX 22-JUN-2001; 2001US-00888224.  
 XX 22-MAY-1996; 96US-00651572.  
 XX 24-APR-1998; 98US-00066544.  
 XX 28-OCT-1999; 99US-00430669.  
 XX (SHOR/) SHORT J M.  
 XX (LAMD/) LAM D E.  
 XX (MATH/) MATHUR E J.  
 XX Short JM, Lam DE, Mathur EJ;  
 XX WPI; 2002-691213/74.  
 XX N-PSDB; ABS54439.  
 XX New isolated nucleic acid encoding a polypeptide having endoglucanase  
 PT activity, useful as a probe for isolating or identifying other  
 PT endoglucanase genes having a sequence that is similar to its sequence.  
 XX Claim 35; Fig 5; 36pp; English.  
 XX The invention discloses an isolated nucleic acid encoding a polypeptide  
 CC from Thermotoga maritima which has an endoglucanase activity. Cellulose  
 CC is a tough, fibrous, water insoluble substance found in the cell walls of  
 CC plants. It consists of an unbranched homopolysaccharide of 10,000 to  
 CC 15,000 D-glucose units in a beta configuration linked by beta 1-4  
 CC glycosidic bonds. The enzymatic hydrolysis of cellulose requires the  
 CC action of both endoglucanases (1,4-beta-D-glucan glucanohydrolase) and  
 CC exoglucanases (1,4-beta-D-glucan cellobiohydrolase). The endoglucanase of



KW emulsion stability; temperature resistance; pseudoplasticity; amylase;  
KW cellulase; extracellular protease; intracellular protease;  
KW glucose dehydrogenase; enzyme.  
XX Xanthomonas campestris.  
XX US2003036176-A1.  
XX 20-FEB-2003.  
XX 10-AUG-2001; 2001US-00927827.  
XX 28-MAR-2001; 2001US-0279493P.  
XX (BOWE/) BOWER S G.  
XX (RAMS/) RAMSEIER T M.  
XX Bower SG, Ramseier TM;  
XX WPI; 2003-625389/59.  
XX N-PSDB; ADD24896.  
XX New transformed cell or organism having reduced or enhanced activity of  
PT at least one protein, useful for producing xanthan gum, which are useful  
PT for providing formulations and properties.  
XX Claim 1; SEQ ID NO 50; 135pp; English.  
XX The present invention relates to polypeptide and polynucleotide sequences  
CC from Xanthomonas campestris which may be used for activity reduction or  
CC enhancement using directed genetic engineering. A transformed cell or  
CC organism having reduced or enhanced activity of at least one such protein  
CC e.g. galactomannanase can be generated by disrupting the gene encoding  
CC the protein. The activity of the protein is reduced by the presence of an  
CC antisense nucleic acid sequence. The nucleic acid sequence of the gene  
CC encoding the protein is a recombinant sequence having at least one  
CC mutation as compared to the wild-type gene encoding the protein. The  
CC transgenic cell or microorganism are useful for producing xanthan gum,  
CC which are useful for providing formulations and properties, such as long-  
CC term suspension and emulsion stability in alkaline, acid, and salt  
CC solutions, temperature resistance, and pseudoplasticity. The present  
CC sequence represents an enzyme relating to the present invention.  
XX Sequence 535 AA;  
SQ  
Query Match 40.1%; Score 796.5; DB 7; Length 535;  
Best Local Similarity 46.6%; Pred. No. 2.7e-66;  
Matches 160; Conservative 61; Mismatches 103; Indels 19; Gaps 9;  
QY 5 YHTSGREILDANNVPRVRIAGINWFGFETCNVYVHGLMSRDRSMLDQIKSLGYNTIRLP 64  
DB 77 YSINSRQIVDDSGKVVQLKGVNVFGFETGNHVMHGLWARKWDMIVQMGLGFNAVLRLP 136  
QY 65 YSDILKPGTMPNSINFRQMNQDQLGTLQVMDKIVAYAGQIGLRILDRHRPDCSGOS 124  
DB 137 FCPATLRSDTMPASIDYSR-NADLQGLTSLQILDKVIAEFNARGMYVLDDHTPDCAGIS 195  
QY 125 ALWYTSVSATWISDLQALQAKYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184  
DB 196 ELWYTSYTYEQLADLRFVANRYKVPYVGLDGLXNEPHGATWGTGNAATDWNKAER 255  
QY 185 AGNAVLNPNLLIFVEGVQ-----SYNGDSYVWGNLQAGQYPPVLANVP-NRLVYSAH 238  
DB 256 GSAALVAVAPKWLIAVEGITDNTFVCTNG-GIFWGNLQPLACTP--LNPANRLLLAP 312  
QY 239 DYATSVYPTQWFDPTFPNMPGINKNMGYLFNQNIAPVWLGFEFTTL--QSTTDQWL 296  
DB 313 VYGPDPVYQSYFNDSPNPNMPAIWERHFGQFAGTH--ALLLGEFGKYGEGDARDKTWQ 370  
QY 297 KTLVQVLRPTAQYACDSFQWTFWSNPDSDGTGILKDDWQTV 339  
DB 371 DALVKVYLR--SKGIN--QGFYWSNPNPDSDGTGILRDWQTSV 408

RESULT 9  
AAW34999  
ID AAW34999 standard; protein; 628 AA.  
XX  
XX AAW34999;  
XX  
XX 27-AUG-2003 (revised)  
XX 21-MAY-1998 (first entry)  
XX  
XX Archaeobacterium AEP11a endoglucanase.  
XX  
XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;  
KW beta-1,4-glycosidic bond; hydrolysis; saccharification;  
KW thermostable enzyme; thermophilic; glycosidase.  
XX  
XX archaeson.  
XX Unidentified.  
XX WO9744361-A1.  
XX  
XX 27-NOV-1997.  
XX  
XX 22-MAY-1997; 97WO-US008793.  
XX  
XX 22-MAY-1996; 96US-00651572.  
XX  
XX (RECO-) RECOMBINANT BIOTOCALYSIS INC.  
XX  
XX Lam DE, Mathur EJ;  
XX  
XX WPI; 1998-018435/02.  
XX N-PSDB; AAT94207.  
XX  
XX Endoglucanase(s), preferably form archaeal bacterium, AEP11a - useful  
PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic  
PT bonds in cellulose.  
XX  
XX Claim 1; Fig 10; 164pp; English.  
XX  
XX This protein comprises an endoglucanase of archaeobacterium AEP11a (Clone  
CC 630P2), a hydrothermal vent isolate. The endoglucanase is capable of  
CC degrading carboxymethylcellulose and of hydrolysing the beta-1,4-  
CC glycosidic bonds in cellulose. It has homology to another endoglucanase  
CC (see AAW34985) of archaeobacterium AEP11a. It can be produced from native  
CC cells or from recombinant host cells, especially prokaryotic host cells  
CC transformed with a plasmid or virus-derived vector including the  
CC endoglucanase DNA (see AAT94207). 24 Endoglucanases (see AAW34986-W35008)  
CC are claimed. They can be used to degrade cellulose for the conversion of  
CC plant biomass into fuels and chemicals, for use in detergents, textiles,  
CC animal feed, waste treatment, and in the fruit juice and brewing  
CC industries for the clarification and extraction of juices. (Updated on 27  
CC -AUG-2003 to correct OS field.)  
XX  
XX Sequence 628 AA;  
SQ

Query Match 38.8%; Score 771; DB 2; Length 628;  
Best Local Similarity 45.5%; Pred. No. 8.8e-64;  
Matches 161; Conservative 50; Mismatches 107; Indels 36; Gaps 11;  
QY 13 ILDANNVPRVRIAGINWFGFETCNVYVHGLMSRDRSMLDQIKSLGYNTIRLPYSDILK 72  
DB 4 VATGEEPIHLFGYNWFGFETPNVYVHGLMSRWNEDMLQIKSLGFNAIRLPCTQSVKP 63  
QY 73 GTMPNSINFRQMNQDQLGTLQVMDKIVAYAGQIGLRILDRHRPDCSGOSALWYTSV 132  
DB 64 GTMPTAIDYAK-NPDLQGLDSVQIMEKIIKAGDLGIFVLLDYHRIKGNTEPLWYTDSP 122  
QY 133 SEATWISDLQALQAKYKGNPTVVGFDLHNEPHDP-----ACWCGDPSIDWRL 180  
DB 123 SEQDYINTWVEVQRFKYNVNVICADLNKPHSSPAPAVTDGSGATWGNNAATDNL 182  
QY 181 AAERAGNAVLNPNLLIFVEGVQ-----SYNGDSYVWGNLQAGQYPPVLANVP- 230

CC activity comprises replacing an active site-associated amino glycosyl-  
CC stabilising amino acid with an amino acid that does not strongly bind a  
CC disaccharide product in the active site. Conversely, the method for  
CC making a glycosyl hydrolase with increased soluble substrate catalytic  
CC activity comprises replacing a hydrophobic substrate-binding amino acid  
CC with a positively charged residue. The invention also discloses mutants  
CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)  
CC produced according to the method of the invention. The Y245G mutant  
CC (AAB48788) has improved activity with insoluble substrates, and the W42R  
CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with  
CC soluble substrates. The invention also encompasses DNA encoding these  
CC mutants. The glycosyl hydrolases of the invention are used as catalysts  
CC for cellulose hydrolysis to produce sugars that can be fermented to  
CC produce fuels such as ethanol. The present sequence represents the  
CC Acidothermus cellulolyticus E1 endoglucanase Y245G mutant  
XX  
SQ Sequence S21 AA;

Query Match 99.1%; Score 1969; DB 4; Length 521;  
Best Local Similarity 99.4%; Pred. No. 2.5e-177;  
Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMDDQIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMDDQIKSLGYNT 60  
QY 61 IRLPYSDILKPGTWPNSINFRMNQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120  
DB 61 IRLPYSDILKPGTWPNSINFRMNQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120  
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
QY 181 AABRAGNAVLSVNPGLIFVEGVQSYNGDSYWGGLQAGQYFVNLNPNRLVYSAHDY 240  
DB 181 AABRAGNAVLSVNPGLIFVEGVQSYNGDSYWGGLQAGQYFVNLNPNRLVYSAHDY 240  
QY 241 ATSVYPTQWTFSDPTFPNNMFGIWNKNGYLFNQNIAPVWLGEFTTLQSTTDDQTLWKLTV 300  
DB 241 ATSVYPTQWTFSDPTFPNNMFGIWNKNGYLFNQNIAPVWLGEFTTLQSTTDDQTLWKLTV 300  
QY 301 QYLRTAQYAGDSFQWTFWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358  
DB 301 QYLRTAQYAGDSFQWTFWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358

RESULT 7  
AAB48786  
ID AAB48786 standard; protein; 521 AA.  
XX  
AC AAB48786;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.  
XX  
KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;  
KW cellulose hydrolysis; ethanol production; fermentation; mutant; mutein.  
XX  
OS Acidothermus cellulolyticus.  
OS Synthetic.  
XX  
FN WO200070031-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 19-MAY-2000; 2000WO-US013971.  
XX  
PR 19-MAY-1999; 99US-0134925P.  
XX  
PA (MIDE ) MIDWEST RES INST.  
XX

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;  
PI Decker SR;  
XX WPI; 2001-061226/07.  
XX  
XX Preparation of glycosyl hydrolase with an increased catalytic activity on  
XX insoluble substrate.  
XX  
XX Claim 16; Page 22-24; 30pp; English.  
XX  
XX The invention relates to a method for making glycosyl hydrolase mutants  
XX with increased catalytic activity with either insoluble or soluble  
XX cellulose substrates relative to the wild-type enzyme. The method for  
XX making a glycosyl hydrolase with increased insoluble substrate catalytic  
XX activity comprises replacing an active site-associated amino glycosyl-  
XX stabilising amino acid with an amino acid that does not strongly bind a  
XX disaccharide product in the active site. Conversely, the method for  
XX making a glycosyl hydrolase with increased soluble substrate catalytic  
XX activity comprises replacing a hydrophobic substrate-binding amino acid  
XX with a positively charged residue. The invention also discloses mutants  
XX of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)  
XX produced according to the method of the invention. The Y245G mutant  
XX (AAB48788) and Y82R (AAB48787) mutants have improved activity with  
XX soluble substrates. The invention also encompasses DNA encoding these  
XX mutants. The glycosyl hydrolases of the invention are used as catalysts  
XX for cellulose hydrolysis to produce sugars that can be fermented to  
XX produce fuels such as ethanol. The present sequence represents the  
XX Acidothermus cellulolyticus E1 endoglucanase W42R mutant  
XX  
SQ Sequence S21 AA;

Query Match 98.9%; Score 1965; DB 4; Length 521;  
Best Local Similarity 99.4%; Pred. No. 6e-177;  
Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMDDQIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSDYRSMDDQIKSLGYNT 60  
QY 61 IRLPYSDILKPGTWPNSINFRMNQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120  
DB 61 IRLPYSDILKPGTWPNSINFRMNQDLQGLTSLQVMDKIVAYAGQIGRLIILDRHRPDC 120  
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
QY 181 AABRAGNAVLSVNPGLIFVEGVQSYNGDSYWGGLQAGQYFVNLNPNRLVYSAHDY 240  
DB 181 AABRAGNAVLSVNPGLIFVEGVQSYNGDSYWGGLQAGQYFVNLNPNRLVYSAHDY 240  
QY 241 ATSVYPTQWTFSDPTFPNNMFGIWNKNGYLFNQNIAPVWLGEFTTLQSTTDDQTLWKLTV 300  
DB 241 ATSVYPTQWTFSDPTFPNNMFGIWNKNGYLFNQNIAPVWLGEFTTLQSTTDDQTLWKLTV 300  
QY 301 QYLRTAQYAGDSFQWTFWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358  
DB 301 QYLRTAQYAGDSFQWTFWNPDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358

RESULT 8  
ADD24922  
ID ADD24922 standard; protein; 535 AA.  
XX  
AC ADD24922;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Xanthomonas campestris cellulase #3.  
XX  
KW Directed genetic engineering; galactomannanase; reduced activity;  
KW enhanced activity; xanthan gum production; suspension stability;  
KW

Db 292 ATSVYPTQWFSFDTFPNNPVGWVFNQNIAPVWLGEFGTTLQSTTDQWLKTLV 341

Qy 301 QYLRTAQYAGDSFQWTFWNSWNPDSGDTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 358

Db 342 QYLRTAQYAGDSFQWTFWNSWNPDSGDTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 399

RESULT 5

AA69508

ID AAY69508 standard; protein; 562 AA.

AC AAY69508;

XX 10-APR-2000 (first entry)

DE Acidothermus cellulolyticus E1 endoglucanase.

XX E1 endoglucanase; cellulose binding domain; CBD; cellulose modification;

XX beta-1,4-endoglucanase; endocellulase; thermostable.

XX Acidothermus cellulolyticus.

OS

XX

XX Location/Qualifiers

FT Peptide 1..41

FT /note= "Putative signal peptide"

FT Peptide 14..41

FT /note= "Putative signal peptide (alternative)."

FT Domain 42..404

FT /note= "Catalytic domain"

FT Region 405..460

FT /note= "Linker region"

FT Domain 461..562

FT /note= "Cellulose binding domain (CBD)"

XX

XX CA2226898-Al.

XX

XX 25-SEP-1999.

XX

XX 25-MAR-1998; 98CA-02226898.

XX

XX 25-MAR-1998; 98CA-02226898.

XX (MIDE ) MIDWEST RES INST.

XX Laymon RA, Adney WS, Thomas SR, Himmel ME;

XX WPI; 2000-087663/08.

DR N-PSDB; AA255924.

XX

XX Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful

PT for labeling or modifying a cellulose and for purifying or immobilizing a

PT binding domain fusion protein to cellulose.

XX

XX Claim 1; Fig 2; 85pp; English.

XX

XX This sequence represents the Acidothermus cellulolyticus E1

CC endoglucanase, which is a beta-1,4-endoglucanase, or endocellulase. The

CC cellulose binding domain (CBD) of E1 endoglucanase, and nucleotides which

CC encode it are specifically claimed. The CBD is believed to be roughly

CC wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge

CC tip is inserted between the microfibrils of the cellulose fibre,

CC disrupting the crystalline structure, and making the cellulose linkages

CC more accessible to the catalytic domain of the E1 endoglucanase. The E1

CC endoglucanase CBD is useful in labelling or modifying the surface of

CC cellulose or other polysaccharides. Such modified cellulose can then be

CC used in textile, pulp, paper, chemical and pharmaceutical industries.

CC CBDs can be used in affinity purification of CBD-fusion proteins, and can

CC also be used to immobilise the CBD-fusion proteins to a cellulose

CC support. CBD-fusion proteins can be used to modify the chemical or

CC physical properties of a cellulose or polysaccharide matrix column and to

CC modify (e.g., roughen or disrupt) a cellulose or polysaccharide fibre.

CC The CBD of the E1 endoglucanase exhibits greater stability at pH 4-8 and

CC has an optimum temperature for stability of 83 degrees Celsius which is

CC not found in CBDs from non-thermophilic organisms

XX Sequence 562 AA;

Query Match 99.4%; Score 1975; DB 3; Length 562;

Best Local Similarity 99.4%; Pred. No. 7.6e-178;

Matches 356; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGGYWHTSGREILDANNVPVRIAGINWGFETCNVYVHGLWSRDYRSLDQIKSLGYNT 60

Db 42 AGGYWHTSGREILDANNVPVRIAGINWGFETCNVYVHGLWSRDYRSLDQIKSLGYNT 101

Qy 61 IRLPYSDDILKPGTMPNSINFQWQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120

Db 102 IRLPYSDDILKPGTMPNSINFQWQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 161

Qy 121 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGGSPSIDWRL 180

Db 162 SGQSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGGSPSIDWRL 221

Qy 181 AABRAGNAVLSVNPENLLIFVEGVQSYNGDSYVWGGNLQAGQYPPVNLVNPRLVYSAHDY 240

Db 222 AABRAGNAVLSVNPENLLIFVEGVQSYNGDSYVWGGNLQAGQYPPVNLVNPRLVYSAHDY 281

Qy 241 ATSVYPTQWFSFDTFPNNPVGWVFNQNIAPVWLGEFGTTLQSTTDQWLKTLV 300

Db 282 ATSVYPTQWFSFDTFPNNPVGWVFNQNIAPVWLGEFGTTLQSTTDQWLKTLV 341

Qy 301 QYLRTAQYAGDSFQWTFWNSWNPDSGDTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 358

Db 342 QYLRTAQYAGDSFQWTFWNSWNPDSGDTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 399

# RESULT 6

AAB48788

ID AAB48788 standard; protein; 521 AA.

XX AAB48788;

XX 09-MAR-2001 (first entry)

XX Acidothermus cellulolyticus E1 endoglucanase mutant, Y245G.

XX E1 endoglucanase; glycosyl hydrolase; insoluble substrate;

XX cellulose hydrolysis; ethanol production; fermentation; mutant; muten.

XX Acidothermus cellulolyticus.

OS Synthetic.

XX WO200070031-Al.

XX 23-NOV-2000.

XX 19-MAY-2000; 2000WO-US013971.

XX 19-MAY-1999; 99US-0134925P.

XX (MIDE ) MIDWEST RES INST.

XX Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;

XX Decker SR;

XX WPI; 2001-061226/07.

XX Preparation of glycosyl hydrolase with an increased catalytic activity on

XX insoluble substrate.

XX Claim 5; Page 27-29; 30pp; English.

XX The invention relates to a method for making glycosyl hydrolase mutants

CC with increased catalytic activity with either insoluble or soluble

CC cellulose substrates relative to the wild-type enzyme. The method for

CC making a glycosyl hydrolase with increased insoluble substrate catalytic

XX PR 15-JUL-1994; 94US-00276213.  
 XX PA (MIDE ) MIDWEST RES INST.  
 XX PI Thomas SR, Laymon RA, Himmel ME;  
 XX DR WPI; 1996-105843/11.  
 XX DR N-PSDB; AAT12337.  
 XX PT New isolated DNA encoding endo:glucanase - obtd from Acidothermus  
 XX PT cellulolyticus, used for prodn of the enzyme for use in cellulose  
 XX PT hydrolysis.  
 XX PS Claim 1; Page 22; 34pp; English.  
 XX CC Acidothermus cellulolyticus El endoglucanase (AAR89927) is useful for  
 XX CC hydrolysing cellulosic biomass to sugars for simultaneous or subsequent  
 XX CC fermentation to ethanol. It shows optimal activity at 83 deg C. The amino  
 XX CC acid sequence of the mature enzyme was deduced from an El endoglucanase  
 XX CC gene (AAT12337) cloned from A. cellulolyticus. Putative signal peptides  
 XX CC were also identified (see also AAR89928-29). Cloning of this gene allows  
 XX CC large-scale, low-cost prodn. of recombinant El endoglucanase, using pref.  
 XX CC Saccharomyces, Zymomonas or E.coli hosts  
 XX SQ Sequence 521 AA;

Query Match 99.6%; Score 1979; DB 2; Length 521;  
 Best Local Similarity 99.7%; Pred. No. 2.8e-178;  
 Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNYVHGLWSDYRSMLDQIKSLGYNT 60  
 DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNYVHGLWSDYRSMLDQIKSLGYNT 60  
 QY 61 IRLPYSDIILKPGTNPNSINFRQNDLQGLTSLQWMDKIVAYAGQIGLRILDRHRPDC 120  
 DB 61 IRLPYSDIILKPGTNPNSINFRQNDLQGLTSLQWMDKIVAYAGQIGLRILDRHRPDC 120  
 QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 QY 181 AAERAGNAVLVSNPNLLIFVEGVQSYNGDSYWWGNNLQAGQYPPVNLVNPRLVYSAHDY 240  
 DB 181 AAERAGNAVLVSNPNLLIFVEGVQSYNGDSYWWGNNLQAGQYPPVNLVNPRLVYSAHDY 240  
 QY 241 ATSVYPTQWTSDFTPFNNMFGIWNKNWGYLFNQNIAPVNLGEGFTTLQSTTDDQTLKTLV 300  
 DB 241 ATSVYPTQWTSDFTPFNNMFGIWNKNWGYLFNQNIAPVNLGEGFTTLQSTTDDQTLKTLV 300  
 QY 301 QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDWQTVTVTKDGLAPIKSSIFDPV 358  
 DB 301 QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDWQTVTVTKDGLAPIKSSIFDPV 358

## RESULT 4

AAU79549  
 ID AAU79549 standard; protein; 562 AA.

XX AC AAU79549;  
 XX DT 24-SEP-2002 (first entry)  
 XX DE A. cellulolyticus cellulase El beta-1,4-endoglucanase precursor.  
 XX KW Gene; ds; transgenic; plant; lignocellulose; cellulase; ligninase;  
 XX KW fermentable sugar; ethanol; fermentation; silage; feed; fuel;  
 XX KW industrial chemical; biodegradation; chloroaromatic;  
 XX KW environmental pollutant; El beta-1,4-endoglucanase; el.  
 XX OS Acidothermus cellulolyticus.

Key Location/Qualifiers  
 1. 41  
 /label= Leader\_sequence  
 42. 562  
 /label= Mature\_cellulase  
 WO200234926-A2.  
 02-MAY-2002.  
 18-OCT-2001; 2001WO-US032538.  
 20-OCT-2000; 2000US-0242408P.  
 (UNMS ) UNIV MICHIGAN STATE.  
 Sticklen MB, Dale BE, Magbool S;  
 WPI; 2002-489947/52.  
 N-PSDB; ABK86729.  
 Producing transgenic plants which after harvest degrade lignin and  
 cellulose to fermentable sugars, by mating transgenic plant comprising  
 DNA encoding cellulase with transgenic plant comprising DNA encoding  
 ligninase.  
 Disclosure; Page 96-99; 126pp; English.  
 The invention discloses the production of a transgenic plant which  
 degrades lignocellulose when the plant is ground. It comprises the  
 production of the transgenic plant including cellulase and ligninase by  
 mating a transgenic plant, containing a DNA encoding a cellulase, and a  
 transgenic plant, containing a DNA encoding a ligninase, where both genes  
 are operably linked to a nucleotide sequence encoding a signal peptide  
 which targets the fusion protein to an organelle of the plant,  
 particularly chloroplasts. The method is useful for producing a  
 transgenic plant (e.g. maize) which degrades lignocellulose when the  
 plant is ground to produce a plant material. This material is useful for  
 converting lignocellulose, in a plant material, to fermentable sugars  
 which are then fermented to ethanol. The transgenic plants also provide a  
 plentiful and inexpensive source of fungal or bacterial cellulases and  
 ligninases which can be used in the production of ethanol. They can also  
 be used for pre-treating silage to increase the energy value of  
 lignocellulosic feeds for cows and other ruminant animals, pre-treating  
 lignocellulosic biomass for fermentative conversion to fuels and  
 industrial chemicals, and biodegradation of chloroaromatic environmental  
 pollutants. The protein sequence presented is the A. cellulolyticus  
 cellulase El beta-1,4-endoglucanase precursor  
 Sequence 562 AA;

Query Match 99.6%; Score 1979; DB 5; Length 562;  
 Best Local Similarity 99.7%; Pred. No. 3.2e-178;  
 Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNYVHGLWSDYRSMLDQIKSLGYNT 60  
 DB 42 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNYVHGLWSDYRSMLDQIKSLGYNT 101  
 QY 61 IRLPYSDIILKPGTNPNSINFRQNDLQGLTSLQWMDKIVAYAGQIGLRILDRHRPDC 120  
 DB 102 IRLPYSDIILKPGTNPNSINFRQNDLQGLTSLQWMDKIVAYAGQIGLRILDRHRPDC 161  
 QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 162 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 221  
 QY 181 AAERAGNAVLVSNPNLLIFVEGVQSYNGDSYWWGNNLQAGQYPPVNLVNPRLVYSAHDY 240  
 DB 222 AAERAGNAVLVSNPNLLIFVEGVQSYNGDSYWWGNNLQAGQYPPVNLVNPRLVYSAHDY 281  
 QY 241 ATSVYPTQWTSDFTPFNNMFGIWNKNWGYLFNQNIAPVNLGEGFTTLQSTTDDQTLKTLV 300

CC (AA848788) has improved activity with insoluble substrates, and the W42R  
CC (AA848786) and Y82R (AA848787) mutants have improved activity with  
CC soluble substrates. The invention also encompasses DNA encoding these  
CC mutants. The glycosyl hydrolases of the invention are used as catalysts  
CC for cellulose hydrolysis to produce sugars that can be fermented to  
CC produce fuels such as ethanol. The present sequence represents the  
CC Acidothermus cellulolyticus E1 endoglucanase Y82R mutant  
XX  
SQ Sequence 521 AA;  
Query Match 100.0%; Score 1986; DB 4; Length 521;  
Best Local Similarity 100.0%; Pred. No. 6.2e-179;  
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMDOIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMDOIKSLGYNT 60  
QY 61 IRLPYSDDLKPGTTPMNSINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120  
DB 61 IRLPYSDDLKPGTTPMNSINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120  
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYWGNGNLOGAGQVPVVLNPNRLVYSAHDY 240  
DB 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYWGNGNLOGAGQVPVVLNPNRLVYSAHDY 240  
QY 241 ATSVYPTQWFSDFTFNNMPCINWKNWGYLFNQNIAPVWLGEFTTLOSTTDQTLWKLTV 300  
DB 241 ATSVYPTQWFSDFTFNNMPCINWKNWGYLFNQNIAPVWLGEFTTLOSTTDQTLWKLTV 300  
QY 301 QYLPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
DB 301 QYLPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
RESULT 2  
AAW39262  
ID AAW39262 standard; protein; 358 AA.  
XX AAW39262;  
XX  
XX 14-MAY-1998 (first entry)  
XX  
XX A. cellulolyticus E1-CAT translated region.  
XX  
XX E1 gene; endoglucanase; E1-CAT; catalytic domain; CD; enzyme; stability;  
XX truncated; cellulose hydrolysis; biomass conversion.  
XX  
XX Acidothermus cellulolyticus.  
XX  
XX US5712142-A.  
XX  
XX 27-JAN-1998.  
XX  
XX 22-FEB-1996; 96US-00604913.  
XX  
XX 26-SEP-1989; 89US-00412434.  
XX 27-JAN-1992; 92US-00826089.  
XX 21-SEP-1993; 93US-00125115.  
XX 15-JUL-1994; 94US-00276213.  
XX (MIDE ) MIDWEST RES INST.  
XX  
XX Chou Y, Himmel ME, Baker JO, Thomas SR, Adney WS;  
XX WPI; 1998-119985/11.  
XX N-PSDB; AAV09659.  
XX  
XX DNA encoding truncated form of E1 endo:glucanase from Acidothermus

PT cellulolytic - and related vectors and transformed cells, expressing  
PT only catalytic domain, is used for biomass conversion and has better heat  
PT stability than complete enzyme.  
XX  
XX Claim 2; Fig 5; 19pp; English.  
XX  
XX This sequence represents a novel Acidothermus cellulolyticus endoglucanase  
XX E1 protein (E1-CAT) which expresses only the catalytic domain (CD) of the  
XX enzyme, without its peptide linker or cellulose binding domain (CBD) and  
XX the truncated enzyme expressed by is used for hydrolysis of cellulose  
XX (biomass conversion). Compared with full-length E1, the truncated enzyme  
XX has better heat stability and higher temperature of maximum activity  
XX  
SQ Sequence 358 AA;  
Query Match 99.6%; Score 1979; DB 2; Length 358;  
Best Local Similarity 99.7%; Pred. No. 1.7e-178;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMDOIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMDOIKSLGYNT 60  
QY 61 IRLPYSDDLKPGTTPMNSINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120  
DB 61 IRLPYSDDLKPGTTPMNSINFRQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120  
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYWGNGNLOGAGQVPVVLNPNRLVYSAHDY 240  
DB 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYWGNGNLOGAGQVPVVLNPNRLVYSAHDY 240  
QY 241 ATSVYPTQWFSDFTFNNMPCINWKNWGYLFNQNIAPVWLGEFTTLOSTTDQTLWKLTV 300  
DB 241 ATSVYPTQWFSDFTFNNMPCINWKNWGYLFNQNIAPVWLGEFTTLOSTTDQTLWKLTV 300  
QY 301 QYLPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
DB 301 QYLPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
RESULT 3  
AAW89927  
ID AAW89927 standard; protein; 521 AA.  
XX AAW89927;  
XX  
XX 08-OCT-1996 (first entry)  
XX  
XX A. cellulolyticus E1 endoglucanase.  
XX  
XX E1 endoglucanase; cellulase; cellulose; cellulose; saccharification; ethanol.  
XX  
XX Acidothermus cellulolyticus.  
XX  
XX Key Location/Qualifiers  
XX Domain 1..363  
XX /label= Catalytic-domain  
XX Domain 364..417  
XX /label= Linker  
XX /note= "proline/serine/threonine-rich linker domain  
XX common to multi-domain microbial cellulases"  
XX Domain 418..521  
XX /label= Cellulose-binding\_domain  
XX  
XX WO9602551-A1.  
XX  
XX 01-FEB-1996.  
XX  
XX 14-JUL-1995; 95WO-US008868.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:04 ; Search time 47.6667 Seconds

(without alignments)  
2122.070 Million cell updates/sec

Title: US-09-997-504A-14

Perfect score: 1986

Sequence: 1 AGGQWHTSGREILDANNVP.....VDTVKDGYLAPIKSIIDFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003as:\*

7: Geneseq2003bs:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1986	100.0	521	4	AAB48787
2	1979	99.6	521	2	AAW3262
3	1979	99.6	521	2	AAW3262
4	1979	99.6	562	5	AAU79549
5	1975	99.4	562	3	RAY69508
6	1969	99.1	521	4	AAW48788
7	1965	98.9	521	4	AAW48788
8	796.5	40.1	535	7	ADD24922
9	771	38.8	628	2	AAW34939
10	771	38.8	841	2	AAW34985
11	770	38.8	553	5	ABG70759
12	762	38.4	430	3	ABG10344
13	762	38.4	458	3	ABG10345
14	688.5	34.7	514	4	AAW96264
15	623	31.4	616	2	AAW13494
16	623	31.4	616	5	AAW13494
17	623	31.4	1426	5	AAW13492
18	623	31.4	1426	5	AAW13492
19	524.5	26.4	425	4	ABW3060
20	272	13.7	508	5	ABW3060
21	271.5	13.7	582	7	ADD24920
22	206.5	10.4	488	5	ABW3510
23	205	10.3	38	2	AAW89930
24	203	10.2	389	5	AAW49427
25	188	9.5	517	6	ABP99336

26	187	9.4	551	5	ABB92442	Herbicida
27	175.5	8.8	762	6	ABP73022	Amino aci
28	173.5	8.7	375	6	ABP73019	Amino aci
29	158.5	8.0	518	4	ABB06928	Micromono
30	154.5	7.8	722	6	ABU24069	Protein e
31	153	7.7	420	6	ABP96833	Candida u
32	146.5	7.4	329	2	AAW35002	Thermotog
33	146.5	7.4	406	2	AAW35002	Thermotog
34	143	7.2	431	3	AAW42172	Arabidops
35	143	7.2	431	3	AAW42172	Arabidops
36	143	7.2	431	5	ABW93972	Herbicida
37	143	7.2	442	3	AAW42171	Arabidops
38	143	7.2	442	3	AAW42171	Arabidops
39	143	7.2	476	3	AAW54123	A mannana
40	143	7.2	490	3	AAW54122	Amino aci
41	142.5	7.2	493	5	ABG71218	Bacillus
42	142	7.2	400	4	AAW09785	Pacibaci
43	141	7.1	335	7	ADD24924	Ralstonia
44	141	7.1	666	2	AAW34992	Thermotog
45	138	6.9	348	4	AAW09788	P. pabuli

## ALIGNMENTS

RESULT 1

AAW48787

ID AAB48787 standard; protein; 521 AA.

XX AC AAB48787;

XX DT 09-MAR-2001 (first entry)

XX DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y82R.

XX KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;

XX KW cellulose hydrolysis; ethanol production; fermentation; mutant; mutain.

XX OS Acidothermus cellulolyticus.

XX OS Synthetic.

XX PN WO200070031-A1.

XX PD 23-NOV-2000.

XX PF 19-MAY-2000; 2000WO-US013971.

XX PR 19-MAY-1999; 99US-0134925P.

XX (MIDE ) MIDWEST RES INST.

XX PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;

XX PI Decker SR;

XX DR WPI; 2001-061226/07.

XX PT Preparation of glycosyl hydrolase with an increased catalytic activity on

XX PT insoluble substrate.

XX PS Claim 17; Page 25-27; 30pp; English.

CC The invention relates to a method for making glycosyl hydrolase mutants with increased catalytic activity with either insoluble or soluble cellulose substrates relative to the wild-type enzyme. The method for making a glycosyl hydrolase with increased insoluble substrate catalytic activity comprises replacing an active site-associated amino glycosyl-stabilising amino acid with an amino acid that does not strongly bind a disaccharide product in the active site. Conversely, the method for making a glycosyl hydrolase with increased soluble substrate catalytic activity comprises replacing a hydrophobic substrate-binding amino acid with a positively charged residue. The invention also discloses mutants of acidothermus cellulolyticus E1 endoglucanase (AAB48786-848788) produced according to the method of the invention. The Y245G mutant

```
QY 223 YPVVL-NVDNRLVYSAHYATSVYQTW-----FSDFTPNNMFGIWNKMGYLFNQNIAP 277
Db 1261 YFINLGQYQKVVYSPHDXGPLYQOPWFYQFTKDTLYND---CWRDNWITMDNGIAP 1317
QY 278 VMLGEFGTTLQSTDTQTLWKLTVQVLRPTAQYGADSFQWTFWMSNPDSGDTGGILKDDWQ 337
Db 1318 LALGEGWGLDGDGNEKWTYLRDVI-----ENHIIHTFCWYNANGSDTGGILGVGYDFS 1371
QY 338 TVDTVKDGYLAP 349
Db 1372 TWDEQYKFLKP 1383

RESULT 14
Q9EYQ0 PRELIMINARY; PRT; 534 AA.
AC Q9EYQ0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Cellulase Cel5-N.
GN CELN.
OS Clostridium cellulolyticum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1521;
RN [1]
RP SEQUENCE FROM N.A.
RA Beilach A., Gaudin C., Gal L.;
RT "A gene cluster of cellulases of Clostridium cellulolyticum.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF316823; AAC45162.1; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000772; P:polyaccharide catabolism; IEA.
DR InterPro; IPRO02105; Dockerin 1.
DR InterPro; IPRO02048; EF-hand.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00404; Dockerin 1; 2.
DR PROSITE; PS00018; EF HAND; 2.
DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.
DR SEQUENCE 534 AA; 59671 MW; 980785F0971E5A59 CRC64;

Query Match 29.6%; Score 587.5; DB 2; Length 534;
Best Local Similarity 32.8%; Pred. No. 3.2e-38;
Matches 133; Conservative 60; Mismatches 144; Indels 69; Gaps 13;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRLMDQIKSLGYNTIRLP 64
Db 38 WLCVGDKIYDMNGREYVLTGANWFGFNCSENVFHG-AMVDVKNILTSVADRIGILLRVP 96
QY 65 YSDIL-----KFG-----TMNSINFYQMNQDLQ-----LTSLQVMDKIVAYAGQI 107
Db 97 ISTELLYSNWTKRPNKVVSVTASNPPYTVNDFYDPATDGPKNSEIFDIIMKYCKEL 156
QY 108 GLRIIDRRHPCSGQS-----ALWY-----TSSVSEATWISDLQALAQRYKGNPTVVGPD 158
Db 157 GIKVMIDVHSPDANNSGHMYPWYGLETTTACMTTIDTKWIDTLTLWLAGYKNDTTLAID 216
QY 159 LNEPH-----DPACWCGDPSIDWRLAERAGNAVLNPNLLIFVEGVQSY- 206
Db 217 LKNEPHKRGYTNAAPTDMAKWNNTDENNKWYAAERCSKEILLAVNPKLLIMEGIEQYP 276
QY 207 -----NGDSY-----WGNLQAGQYFVVLNPN-RLVYSAHYATSV 244
Db 277 KTEGYTFTDPVWVGASGDAAPWGGWGNLGRVKDYPIDLGLPLNSQIVYSPHDYGPSV 336
QY 245 YPQTWF-SDFTFNNMFGIWNKMGYLFNQNIAPVWLGEFTTLQSTDTQTLWKLTVQYL 303
Db 337 YNOSWFDKFTTQTLDDYWDYTDWAYIDDQKIAPLLIGEGWGGFMDGAKNKKWTLRLDM 396
```

```
QY 304 RPTAQYGADSFQWTFWMSNPDSGDTGGILKDDWQTVDTVKDGYLAP 349
Db 397 I-----KNRINHTFWCLNPNSGDTGELIGNDWTWDEEYGLLKP 436

RESULT 15
Q7X2N2 PRELIMINARY; PRT; 616 AA.
AC Q7X2N2;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Endoglucanase.
GN CEL5B.
OS Thermomonospora fusca.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.
OX NCBI_TaxID=2021;
RN [1]
RP SEQUENCE FROM N.A.
RA Posta K., Beki E., Kukolya J., Hornok L.;
RT "Phylogenetic relationships of Tf cel5B, a new endoglucanase encoding
gene from Thermobifida fusca.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY298814; AAP56348.1; -.
DR SEQUENCE 616 AA; 67701 MW; 24FFC1EA1A3F5639 CRC64;

Query Match 29.0%; Score 574; DB 2; Length 616;
Best Local Similarity 34.0%; Pred. No. 4.6e-37;
Matches 129; Conservative 63; Mismatches 149; Indels 38; Gaps 13;

QY 1 AGGGYW-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRLMDQIKSLGYN 59
Db 33 SGTADWLHTDGNRIVDSAGNEVLTGANWFGFNTSERMFHGLWAANIEDITSMAAERGIN 92
QY 60 TIRLPYSDIL---KPGTM-PNSINFYQMNQDLQSLQVMDKIVAYAGQIGLRIILDR 115
Db 93 MVRVFIQTQLLEWKNQAGSGVNEY-VNPELAGMNTLEVDYWLQLCIEYGLKWLMDV 151
QY 116 H--RPDCSGQ--SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----166
Db 152 HSAEADNSGHYYPVYKGDITTEDFYTAWEVWTERYKXNDTIVAADIKNEPHGKANETPR 211
QY 167 ACWCGDPSID-WRLAERAGNAVLNPNLLIFVEGVQSYNGD-----SYW 212
Db 212 AKWD-GSTDIDNFKHVCETAGKRILAINPNMLILCEGIETPKDQDQWSSTDGRDYISTW 270
QY 213 WGNLQAGQYFVVLNVPVRIAGINWFGFETCNVYVHGLSRDYSRLMDQIKSLGYL 270
Db 271 WGNLQAGQYFVVLNVPVRIAGINWFGFETCNVYVHGLSRDYSRLMDQIKSLGYL 330
QY 271 FNQNIAPVWLGEFTTLQSTDTQTLWKLTVQVLRPTAQYGADSFQWTFWMSNPDSGDTGG 330
Db 331 HEDDIAPIALLIGEWGGLDGDGNEKWTYLRDVI-----EKMHHTFWALNPNSGDTGG 384
QY 331 ILKDDWQTVDTVKDGYLAP 349
Db 385 LLNYDWTWDEAKYFLKP 403

Search completed: August 2, 2004, 16:36:57
Job time : 37.3333 secs
```

```
QY 338 TVDTVKDGYLAP 349
DB 554 TVDEKVAIIVP 565

RESULT 12
ID O24820 PRELIMINARY; PRT: 1000 AA.
AC O24820;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Beta-glucanase.
OS thermophilic anaerobe NA10.
OC Bacteria.
OX NCBI_TaxID=67756;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NA10;
RA Miyake K., Machida Y., Hattori K., Iijima S.;
RT "Characterization of a multi-domain cellulase from an extremely
thermophilic anaerobe strain NA10.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
DB EMBL; AB008029; BAA22939.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR Pfam; PF001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDLASE10.
DR PRODOM; PD001947; CBD_3; 1.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 31.5%; Score 623.5; DB 2; Length 1000;
Best Local Similarity 35.5%; Pred. No. 1.1e-40;
Matches 133; Conservative 69; Mismatches 134; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNTRLRP 64
DB 590 WLYVSGNKIVDKGRPWLTVGNWFGYNTGTVFDGVWSCNLKSLAEIANRGFNLLRVP 649
QY 65 YSDIL---KPGTWP-N-SINFYQVQNDLQGLTSQVMDKIVAYAGQIGLRIILDRH--RP 118
DB 650 ISASLIILNWSQGIYPKPNINYY-VNPELEGNLSLEVDIVVQTKCEVGLKIMLDIHSIKT 708
QY 119 DCSQ-SALWYTSVSEATWISDQALAQRYKGNPTVVGFDLNEPHDP-----ACWG 170
DB 709 DAMGHIPVWVDEKFTPEDFYKACEWITNRYKNDTIIAIDLKNEPHGKFWQDTTFAKD 768
QY 171 CGDPSIDWRLAERAGNAVLNVNPNLLIFVGVQSYNGD-----SYWGGNLL 217
DB 769 NSTDINNWKYAAETCAKRIILNPNLLIVIGIEIAYPKDDVTWTSKSSDYSTWGGNLL 828
QY 218 QGAGQYFVWL-NVFNRLVYSAHDYATSVYPTWFSDDTFPNN--MPCGIWKNNGYILFNQ 274
DB 829 RVKRYPIINLQYKQVYSPHDYGPVYQOPWFY-PGFYKESLLQDCWFPNWAYIMEEN 887
QY 275 IAPVWLGEFTTLOSTDQTLKTLVQVLRPTAQYAGDSFQWTFWSWNPDSGDTGGLK 334
DB 275 IAPVWLGEFTTLOSTDQTLKTLVQVLRPTAQYAGDSFQWTFWSWNPDSGDTGGLK 334

888 IAPLLIGEWGGLDGNADNEKMKYLRDYII-----ENHIIHTFCWFNANGDTGTGLVGY 941
335 DWQTVDTVKDGYLAP 349
942 DFTTWDEKKYSFLKP 956

RESULT 13
ID Q9X3P6 PRELIMINARY; PRT: 1426 AA.
AC Q9X3P6;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CelB.
OS Caldicellulosiruptor sp. Tok7B.1.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;
OC Caldicellulosiruptor.
OX NCBI_TaxID=80339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tok7B.1;
RX MEDLINE=20171169; PubMed=10706665;
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,
RA Bergquist P.L.;
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme
thermophile Caldicellulosiruptor isolate Tok7B.1.";
RL Curt. Microbiol. 40:333-340(2000).
CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY F (FAMILY 10 OF GLYCOSYL
HYDROLASES).
DB EMBL; AF078737; AAD30364.1; -.
DR HSSP; Q06851; INBC.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001589; Actbind_actnin.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001000; Glyco_hydro_10.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM_3; 3.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00331; Glyco_hydro_10; 1.
DR PRINTS; PR00134; GLHYDLASE10.
DR PRODOM; PD001947; CBD_3; 3.
DR SMART; SM00633; Glyco_10; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;

Query Match 31.1%; Score 617; DB 2; Length 1426;
Best Local Similarity 35.5%; Pred. No. 5.7e-40;
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY 10 GREILDANNVPRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNTRLRPYSDDI 69
DB 1022 GNKIVDKGRPWLTVGNWFGYNTGTVFDGVWSCNLKSLAEIANRGFNLLRVPISAE 1081
QY 70 L---KPGTWP-N-SINFYQVQNDLQGLTSQVMDKIVAYAGQIGLRIILDRH--RPDCSQ 123
DB 1082 ILNWSKGIYPKPNINYY-VNPELEGLTSLEVDFVFKTCKEVGLKIMLDIHSAKTDAMGH 1140
QY 124 -SALWYTSVSEATWISDQALAQRYKGNPTVVGFDLNEPHDP-----ACWCCGDP 175
DB 1141 IYPVWYTDITPEDYKACEWITERYKNDTIIAIDLKNEPHGKFWQDSVFAKNDNSTDI 1200
QY 176 IDWRLAERAGNAVLNVNPNLLIFVGVQSYNGD-----SYWGGNLLQAGQ 222
DB 1201 NNWKYAAETCAKRIILNPNLLIVIGIEIAYPKDDVTWTSKSSDYSTWGGNLLRGYK 1260
```



```
QY 338 TVDTKVGGLAP 349
Db 377 TIDTKLALVQP 388

RESULT 10
Q87AH4 PRELIMINARY; PRT; 614 AA.
AC O87AH4
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase.
GN ENGXA OR PD1851.
OS Xylella fastidiosa (strain Temecula / ATCC 700964).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=183190;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=22421331; PubMed=12533478;
RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA Goldmann M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,
RA Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA Coutinho L.B., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA Baia G.S., Blanco S.R., Brito M.S., Cannavan P.S., Celestino A.V.,
RA da Cunha A.F., Fenille R.C., Brito J.A., Formighieri E.F., Kishi L.T.,
RA Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sasaki F.T., Sena J.A.D.,
RA de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA Civerolo E.D., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA Kitajima J.P.;
RT "Comparative analyses of the complete genome sequences of Pierce's
RT disease and citrus variegated chlorosis strains of Xylella
RT fastidiosa".
RL J. Bacteriol. 185:1018-1026 (2003).
DR EMBL; A0012560; AAC29683.1; -.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002952; Egshe11.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PRINTS; PR01228; EGGSHELL.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 614 AA; 62041 MW; 29EC67B42F4BC3F7 CRC64;

Query Match 32.6%; Score 645.5; DB 16; Length 614;
Best Local Similarity 39.9%; Pred. No. 9.6e-43;
Matches 141; Conservative 59; Mismatches 125; Indels 29; Gaps 12;

QY 5 YHSTGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMGLDQIKSLGYNTIRLP 64
Db YSI:SHGKVVDDKGN-QIQLRGVNWFGFETGDHVTVGLMARNWKEFITQLQGNGFNAIRLP 84
QY YSSDILKPGTMMNSINFYQNNQDLGLTSLOVMDKIVAYAGQIGLRITLDRH-R 124
Db YSI:SHGKVVDDKGN-QIQLRGVNWFGFETGDHVTVGLMARNWKEFITQLQGNGFNAIRLP 84
QY FCPANLNSNTSPSSID-YSNPNPGLGLSLQLDKVVKLSLRDMYVLLDHRPDCSAIS 143
Db FCPANLNSNTSPSSID-YSNPNPGLGLSLQLDKVVKLSLRDMYVLLDHRPDCSAIS 143
QY ALWYTSVSEATWISDLQALAQRYKGNPTVGVDFLNEPHDPCWCGGPPS:DWRLAAR 184
Db ELMTYSYSEKQWIDDLRFVARYANVGHVGLGVNNEPHGRTTWGTPKTDWNTAVEH 203
QY AGNAVLSVNPNTLIFVEGVQSYNGDS-----YWGNGNLQAGQYPPVVLNVP-NRLVY 235
Db AAAAILEAAPKWLIGVEGI-----GENPICSSTTICHFNGENLEPMDCTP--LKVPAHLL 257

QY 236 SAHDYATSVYPTQWFSDDPTFPNNMGPWKNKNGYLFNQNTAPVWLGEFGTTL--QSTTDD 293
Db MPHVGPDVYVQPYFNSPDPFNMAALWDKHFHFAKAGYA-MAIGFEGKYGEGPRDI 316
QY 294 TWLKTLYOYLPRPTAQYG-ADSFQWTFWSNPDSDTGGILKDDQWTV--DTVK 343
Db AMQNAFVDYL---ISIGVTDAF---YWAANQNSVDITGGWGVNDWTTTPRDDKVK 363

RESULT 11
Q9L3J2 PRELIMINARY; PRT; 660 AA.
AC Q9L3J2
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 1,4-beta-cellobiohydrolase.
GN CELO.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1515;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=F7;
RA Zverlov V.V., Schwarz W.H.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ275975; CAB76938.1; -.
DR HSP; P54583; 1FCE.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO; GO:0016787; F:hydrolyase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0000272; P:polyaccharide catabolism; IEA.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR002105; Dockerin_1.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR000437; Prok_lipoprot_S.
DR Pfam; PF00150; cellulase; 1.
DR Pfam; PF00404; Dockerin_1; 2.
DR PROSITE; PS00448; CLOS_CELLULOsome_RPT; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hydrolyase.
SQ SEQUENCE 660 AA; 75247 MW; 04A042002C288CFC CRC64;

Query Match 32.0%; Score 634.5; DB 2; Length 660;
Best Local Similarity 36.0%; Pred. No. 7.9e-42;
Matches 134; Conservative 66; Mismatches 137; Indels 35; Gaps 10;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCN-YVHGLRSDYRSMGLDQIKSLGYNTIRL 63
Db WLHVEGNLIKDAQGNVTYLTGINWFGFETDGANGFGLNKNLESDSLDLMAKLGFTLRI 260
QY 64 PYSDDIL---KPGTMMNS--INFYQNNQDLGLTSLOVMDKIVAYAGQIGLRITLDRH-R 117
Db PISAEIILQWNGRVERTSFVNTYE-NPRLDGLSLLEILYTNHMKNGKMAIDHSS 319
QY 118 PDCSQGALWYTSVSEATWISDLQALAQRYKGNPTVGVDFLNEPH-----DPAWGC 171
Db TKDSYQENLWYKDI:TWEEFTFAWKWIVERYKDDDTVIAVDLKNPHGKYSGNIAKWD 379
QY 172 GDPIDWELAAERAGNAVLSVNPNTLIFVEGVQSYNGDSY-----YWGNGNLQ 218
Db SNDPNMWRKAAEITABEILAINPNLLIVVEGVAYPMVEGYDYNCGEFTTYCNWNGNLR 439
QY 219 GAGQYPPVVLNVPNTLIFVEGVQSYNGDSY-----YWGNGNLQAGQYPPVVLNVP 277
Db GVAHPVVISAPDKLVISVHDYGPDIYMQPWFKKDFDINTLYECWYPNWYIYEQNIAP 499
QY 278 VWLGEFGTTLQSTDDQWTLKTVQYLRPTAQYGADSFQWTFWSNPDSDTGGILKDDQW 337
Db MLIGEWGGLINENNRKWLCELATFI-----ASKLHHTFAFNPSADTGGGLNLEBDK 553
```

[illegible]

```
Db 262 YGPDVFQSYFNDSNFPNNMPAIDRHFQGFAGSH--ALLLGBFGKYGEGDARDKVKWD 319
Qy 298 TLVQYLRPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 320 ALVKYLR---SKGIN--EGFYWSNPNDSGDTGGILRDDWTSV 356

RESULT 6
O58925 PRELIMINARY; PRT; 458 AA.
AC O58925
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 458AA long hypothetical endo-1,4-beta-glucanase.
GN PH1171.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN SEQUENCE FROM N.A.
RP STRAIN=OT3;
RC MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo H. Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.,
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5155-76 (1998).
DR EMBL; AP000005; BAA30271.1; -.
DR PIR; E71059; E71059.
DR HSSP; P54583; 1BCE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco_hydro_5.
DR InterPro; IPR006162; Ppantne_S.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR PROSITE; PS00012; GHOSPHOPANTETHEINE; 1.
KW Complete proteome.
SQ SEQUENCE 458 AA; 51930 MW; CAE48AD43A8EB654 CRC64;

Query Match 37.9%; Score 751; DB 17; Length 458;
Best Local Similarity 45.0%; Pred. No. 2.6e-51;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTIRLPYSD 67
Db 54 TSGBE-----TPIHLFGVNWFGFETPNVHGLWKRNMEDMLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTMNSINFYQNDQLGSLTSLOVMDKIYAVAGQIGLRIILDRHRPCSGSALW 127
Db 108 ESKVPGTQPTGID-YSKNPDLRLGSLQIMEKIKKAGDLGIFVLLDYHRICTHIEPLW 166
Qy 128 YTSVSSEATWISDLQALAQRYKGNPTVGFDLHNEPHD-----PACWGCSDPS 175
Db 167 YTEDFSEDFINWIEVAKFGKXVNVIGADLNKNEHSVTSPPAAYTDGTGATWGNFA 226
Qy 176 IDWRLAERAGNAVLSNPNLLIFVEGVQSYN-----GDSYWWGNLQAGQYPVV 226
Db 227 TDWNLAAERIGKAILKVPAPHLIFVEGTQFTNPKTDSYKWKGNWGNLNAVKDYPV- 285
Qy 227 LNVP-NRLVVSADHYATSVVPQWFSDEPT-FPNMPPGINWKNWGLFNQNIAPVWLGBFG 284
Db 286 -NLPRNLKVLISPHVYGPDVNTQYFGPAKFPNLDPIWTHHGYKLELGSYVIVGBFG 344
Qy 285 TTL---QSTTDQTLKTLVOYLRLPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 345 GKYGHGGPRDVIWQNKLVDMW--IENKFCDFE---YWSNPNDSGDTGGILQDDWTTI 397
```

```
RESULT 7
Q9V052 PRELIMINARY; PRT; 514 AA.
AC Q9V052
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Endoglucanase.
GN CELB-LIKE OR PYRAB09400 OR PAB0632.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN SEQUENCE FROM N.A.
RP STRAIN=GES / Orsay;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248285; CAB49854.1; -.
DR PIR; E75142; E75142.
DR HSSP; P54583; 1BCE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 514 AA; 59980 MW; CFP80CAB1525337C5 CRC64;

Query Match 35.0%; Score 694.5; DB 17; Length 514;
Best Local Similarity 41.9%; Pred. No. 9.6e-47;
Matches 149; Conservative 57; Mismatches 121; Indels 29; Gaps 10;

Qy 7 HTSGREILDANNV-----PVRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNT 60
Db 30 YTAENGIIFVQNVTTGKPKLYLHGVSFWGFKLKHVVYGLDKRNWKDILKDKVRLGFNA 89
Qy 61 IRLPYSDILKPGTMN--SINFYQNDQLGSLTSLOVMDKIYAVAGQIGLRIILDRHRP 118
Db 90 IRLPFCSESRPDRTPSPERIN-YELNPDNLKLTSLNEKILIEYANSIGLYILLDYHRI 148
Qy 119 DCSGQSALWYTSVSSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPAWCGSDPSIDW 178
Db 149 GCEETELPWTENYSEEQYIKDWIFLAKRFGKYPNVIGADIKNEPHGEGAGWTGDER-DF 207
Qy 179 RLAAERAGNAVLSNPNLLIFVEGVQ-----SYNGDSYW--WGNLQAGQYPVWL 227
Db 208 RLFAEKVGREILKVPAPHLIFVEGTQYTHVNIIDEIIEKKGWMTFWGNLWGVKDYVRL 267
Qy 228 NVPNRLVVSADHYATSVVPQWFSDEPTFPNNMPPGINWKNWGLFNQNIAPVWLGEFGTTL 287
Db 268 -PRGVVYSPHVYGSPVYMDYFKSPDPFPNNWPIIWEHFGYLDLNLV-LV-IGEWGNY 325
Qy 288 QSTTDQTLKTLVOYLRLPTAQYAGDSFQWTFWSNPNDSGDTGGILKDDWQTVTVK 343
Db 326 EG-LDKVWQDAFVKWLKKIY-----NFFYWCNLNPSGDTGGIFLDDWKTVMWEK 375

RESULT 8
Q9PF60 PRELIMINARY; PRT; 592 AA.
AC Q9PF60
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase.
GN XF0818.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
```

```

Db 165 SELWTSYEPESRWISDRMLAKRYASDPTVIGVDLHNEPHGAATWGTGAATTDWRAAAE 224
Qy 184 RAGNAVLSVNPENLLIFVEGV-QSYNGDSYWGGLQAGAGQPVVIANVPLVYSAHDYAT 242
Db 225 RGNNAVLAENPKLLVLVEGDHQAQDGTGTWGGALDSAAATASVRLTVANRVVTSRPHDYS 284
Qy 243 SVYPTQWFSOPTFNNMPEGNKWNKGNLGNFNAPVLMGFEFGTTLQSTDTQTLWKLTVQY 302
Db 285 TIYQCPWFASNYTNLPGLIWDHAGYLAAXDAPVLVGEFGTKLEFASDKQMLNTLVGY 344
Qy 303 LRPTAQYGAASFOQWFSWNPDSGDTGGILKDDWQTVDTVDKGYLAPI 350
Db 345 LSST---GISS---SFAFNPDSDGTGGIVKSDWVTPEQAKLDALAPI 386

RESULT 4
Q9EVR6 PRELIMINARY; PRT; 518 AA.
AC Q9EVR6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-OCT-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Extracellular endoglucanase (ENGXCA protein) precursor.
GN ENGXCA.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRLL-B1459;
RA Schroeter K., Puchler A., Becker A.;
RT "engXCA major extracellular endoglucanase.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304415; CACI8529.1; -.
DR HSSP; P54583; 18CE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008965; Cellul bind.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 518 EXTRACELLULAR ENDOGLUCANASE (ENGXCA
FT PROTEIN).
SQ SEQUENCE 518 AA; 55588 MW; 9A0FF678E3BF712C CRC64;

Query Match 40.0%; Score 792.5; DB 2; Length 518;
Best Local Similarity 46.9%; Pred. No. 1.6e-54;
Matches 161; Conservative 59; Mismatches 104; Indels 19; Gaps 9;

Qy 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDDQIKSLGYNTIRLP 64
Db 26 YSINNSRQIVDDSGKVLKGVNFGFETGNVHGLWARNWKEMIVQMQLGFLNAVRLP 85
Qy 65 YSDILKPGTNPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDCSGQS 124
Db 86 FCPATLRSDTTPASID-YSRNADLQGLTSLQILDKVIAEFNARGMYVLLDHHPTDCAGIS 144
Qy 125 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFPLHNEPHDPACWCGDPSIDWRJAAER 184
Db 145 ELWYTSYEAQWLADLRFVANRYKNVPYVGLDLKNEPHGAATWGTGNAATDWNKAAER 204
Qy 185 AGNAVLSVNPENLLIFVEGVQ-----SYNGDSYWGGLNQGAGQPVVIANVPLVYSAH 238
Db 205 GSAVLAVAPKWLIAVEGITDNPVCSITNG-GIFGGLNQLPLACTP--LNIPANRLLLAPH 261
Qy 239 DYATSVPTQWFSDFPTFNNMPEGNKWNKGNLGNFNAPVLMGFEFGTTL--QSTDTQTL 296
Db 262 VYGPDPVFQSYFNSFNPNMPEALWERHFGQFAGTH--ALLIGFGGKYGEGARDKTMQ 319
Qy 297 KTLVQYLRPTAQYGAADSEQWTFNSWNPDSGDTGGILKDDWQTV 339

```

```

Db 320 DALVKYLR---SKGIN--QGFVSWNPNSGDTGGILRDDWTSV 357

RESULT 5
Q8PPS3 PRELIMINARY; PRT; 474 AA.
AC Q8PPS3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase.
GN ENGXCA OR XAC0612.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lenos M.V.F.,
RA Locali B.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.N., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Sperinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:453-483 (2002).
DR EMBL; A5011689; AAM35501.1; -.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 51294 MW; 379B84D63CASA31B CRC64;

Query Match 39.9%; Score 791.5; DB 16; Length 474;
Best Local Similarity 46.2%; Pred. No. 1.7e-54;
Matches 158; Conservative 61; Mismatches 104; Indels 19; Gaps 9;

Qy 6 WHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDDQIKSLGYNTIRLP 65
Db 26 YSINNSRQIVDDSGKVLKGVNFGFETGNVHGLWARNWKEMIVQMQLGFLNAVRLP 85
Qy 66 YSDILKPGTNPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDCSGQS 125
Db 86 FCPATLRSDTTPASID-YSRNADLQGLTSLQILDKVIAEFNARGMYVLLDHHPTDCAGIS 144
Qy 126 ALWYTSVSEATWISDLQALAQRYKGNPTVVGFPLHNEPHDPACWCGDPSIDWRJAAER 185
Db 145 ELWYTSYEAQWLADLRFVANRYKNVPYVGLDLKNEPHGAATWGTGNAATDWNKAAER 204
Qy 186 AGNAVLSVNPENLLIFVEGVQ-----SYNGDSYWGGLNQGAGQPVVIANVPLVYSAH 239
Db 205 GSAVLAVAPKWLIAVEGITDNPVCSITNG-GIFGGLNQLPLACTP--LNIPANRLLLAPH 261
Qy 240 YATSVPTQWFSDFPTFNNMPEGNKWNKGNLGNFNAPVLMGFEFGTTL--QSTDTQTL 297

```

```
QY 124 SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
DB 156 SELWYTSQYSPESWISDWRMLAKRYAAEPTVIGVLLHNEPHGAATWGTGAATDWRRAAE 215
QY 184 RAGNAVLSVNPENLLIFVEGV---SYNGDSYWMGNLQAGQYPPVVLNVPNRLVYSAHYD 240
DB 216 RAGNAVLSVNPENLLIFVEGV---SYNGDSYWMGNLQAGQYPPVVLNVPNRLVYSAHYD 275
QY 241 ATSVPTQWFSPTFPNNPNNPGIWNKNWGYLFNONTAPVWLGEFG--TTLQSTDTQWTKT 298
DB 276 GFCVSSQPFNDSTFESNLPFAWDQWGVYSKQNTAPVLVGEFGGRNVLDLSPESGKQWNA 335
QY 299 LVQYLRTAQYCADSFQWTFWGNPDSGDTGILKDDWQTVTVXDGVLAPKSSIFDPV 358
DB 336 LVDYI-----GANNLYFYWNLNPNNSGDTGILLDDWTTWNRPKQDML----SRIMKPV 385

RESULT 2
Q9K5C7 PRELIMINARY; PRT; 745 AA.
ID Q9K5C7
AC Q9K5C7
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cellulase precursor (Fragment).
GN CELA.
OS Clavibacter michiganensis.
OG Plasmid pCML
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=28447;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Melezus D.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-614 FROM N.A.
RC STRAIN=NCPPB 382;
RX MEDLINE=20331586; PubMed=10875331;
RA Jahr H., Dreier J., Melezus D., Bahro R., Eichenlaub R.J.;
RT "The endo-beta-1,4-glucanase CelA of Clavibacter michiganensis subsp.
RT michiganensis is a pathogenic determinant required for induction of
RT bacterial wilt of tomato.";
RL Mol. Plant Microbe Interact. 13:703-714 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Jahr H.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
DR EMBL; X62582; CAA44467.2; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Plasmid; Signal.
FT SIGNAL 1 66 POTENTIAL.
FT CHAIN 67 >745 CELLULASE.
FT NON_TER 745 745
SQ SEQUENCE 745 AA; 77986 MW; 5B45015EB8D4F27C CRC64;

Query Match 46.1%; Score 913.5; DB 2; Length 745;
Best Local Similarity 50.3%; Pred. No. 6.4e-64;
Matches 175; Conservative 51; Mismatches 111; Indels 11; Gaps 4;
```

```
QY 4 GYHWTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTIRL 63
DB 70 GWLHTAGKIVTAGAPYITIRGIAWFMETSSCAPHGLDTITLACGMQHKKMGFTTVRL 129
QY 64 PYSDDLKPGTNPNSINFYQMNQDLOGLTSLQWMDKIVAVAGQIGLRILDRHRPDSGQ 123
DB 130 PFSNQCL----AASGVTGVSADPSLAGLTPLOQVMDHVWASAKSAGLDVILDQHRPDSGQ 185
QY 124 SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
DB 186 SELWYTSQYSPESWISDWRMLAKRYAAEPTVIGVLLHNEPHGAATWGTGAATDWRRAAE 245
QY 184 RAGNAVLSVNPENLLIFVEGV-QSYNGDSYWMGNLQAGQYPPVVLNVPNRLVYSAHYD 242
DB 246 RGNNAVLAENPNLLVLVEGIDHEADGSGTWMGCGALGLVGNAPVRLSVANRVVYSPHDPS 305
QY 243 SVYPTQWFSPTFPNNPNNPGIWNKNWGYLFNONTAPVWLGEFGTTLQSTDTQWTKTVQY 302
DB 306 TIYQGSWFSASNYPANLPGLIWDHGWGLAKKDLAPVLVGEFGTKFETTSKQWNLTVGY 365
QY 303 LRPTAQYCADSFQWTFWGNPDSGDTGILKDDWQTVTVXDGVLAP 350
DB 366 LSSI---GISS---SFWAFNPNNSGDTGIVKSDWVTEQA KLDALAPI 407

RESULT 3
Q9AF65 PRELIMINARY; PRT; 727 AA.
ID Q9AF65
AC Q9AF65;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Cellulase CelA.
OS Clavibacter michiganensis (subsp. sepedonicus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=31964;
RN [1]
RP SEQUENCE FROM N.A.
RA Laine M., Haapalainen M., Wahlroos T., Kankare K., Nissinen R.,
RA Kassuwi S., Mettler M.C.;
RT "The cellulase encoded by the native plasmid of Clavibacter
RT michiganensis subsp. sepedonicus plays a role in virulence and
RT contains an expansin-like domain.";
RL Physiol. Mol. Plant Pathol. 0:0-0 (2001).
DR EMBL; AY007311; AAK16222.1; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM 2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
SQ SEQUENCE 727 AA; 75811 MW; 96DF9B664873985D CRC64;

Query Match 45.7%; Score 906.5; DB 2; Length 727;
Best Local Similarity 50.3%; Pred. No. 2.2e-63;
Matches 175; Conservative 48; Mismatches 114; Indels 11; Gaps 4;
```

```
QY 4 GYHWTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRMLDQIKSLGYNTIRL 63
DB 49 GWLHTAGKIVTAGAPYITIRGIAWFMETSSCAPHGLDTITLACGMQHKKMGFTTVRL 108
QY 64 PYSDDLKPGTNPNSINFYQMNQDLOGLTSLQWMDKIVAVAGQIGLRILDRHRPDSGQ 123
DB 109 PFSNQCL----AASGVTGVSADPSLAGLTPLOQVMDHVWASAKSAGLDVILDQHRPDSGQ 164
QY 124 SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:29 / Search time 36.3333 Seconds  
(without alignments)  
3108.867 Million cell updates/sec

Title: US-09-997-504A-12  
Perfect score: 1982  
Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25:\*

- 1: sp\_archea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp Vertebrate:\*
- 14: sp Unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	1044	52.7	397	2 Q8RP23	Q8RP23 paenibacill
2	913.5	46.1	745	2 Q9K5C7	Q9K5C7 clavibacter
3	906.5	45.7	727	2 Q9AF65	Q9AF65 clavibacter
4	792.5	40.0	518	2 Q9EV76	Q9EV76 xanthomonas
5	791.5	39.9	474	16 Q8P8S3	Q8P8S3 xanthomonas
6	751	37.9	458	17 Q58325	Q58325 pyrococcus
7	694.5	35.0	514	17 Q9V052	Q9V052 pyrococcus
8	683.5	34.5	592	16 Q9PF60	Q9PF60 xylella fas
9	674	34.0	482	16 Q97KK6	Q97KK6 clostridium
10	645.5	32.6	614	16 Q87AH4	Q87AH4 xylella fas
11	634.5	32.0	660	2 Q9L3J2	Q9L3J2 clostridium
12	623.5	31.5	1000	2 Q24820	Q24820 thermophil
13	617	31.1	1426	2 Q9X3B6	Q9X3B6 caldicellul
14	587.5	29.6	534	2 Q9EYQ0	Q9EYQ0 clostridm
15	574	29.0	616	2 Q7X2N2	Q7X2N2 thermomonos
16	509.5	25.7	425	2 Q93Q07	Q93Q07 thermus cal

17	498	25.1	341	2	P96310	P96310 anaerocellu
18	485	24.5	574	2	Q8S3V3	Q8S3V3 cellulomona
19	398	20.1	630	2	Q8RJY7	Q8RJY7 sigmatella
20	328	16.5	565	16	Q9PF68	Q9PF68 xylella fas
21	317.5	16.0	569	16	Q87AG9	Q87AG9 xylella fas
22	281.5	14.2	590	16	Q8P513	Q8P513 xanthomonas
23	267	13.5	508	10	Q9LTM8	Q9LTM8 arabidopsis
24	252.5	12.7	522	10	Q8SAB6	Q8SAB6 arabidopsis
25	214	10.8	526	10	Q9LFS2	Q9LFS2 arabidopsis
26	204.5	10.3	488	10	Q9LFE7	Q9LFE7 arabidopsis
27	194.5	9.8	354	16	Q8PEF2	Q8PEF2 xanthomonas
28	190.5	9.6	561	10	Q9LTM0	Q9LTM0 arabidopsis
29	189.5	9.6	555	10	Q7XU04	Q7XU04 oryza sativ
30	181.5	9.2	437	3	Q99036	Q99036 trichoderma
31	180.5	9.1	694	2	Q9Z187	Q9Z187 bacillus st
32	180	9.1	439	3	Q8Z401	Q8Z401 agaricus bi
33	179	9.0	439	3	Q9P893	Q9P893 agaricus bi
34	178.5	9.0	377	3	Q00012	Q00012 aspergillus
35	168.5	8.5	558	10	Q7XU06	Q7XU06 oryza sativ
36	166.5	8.4	357	16	Q8PRD3	Q8PRD3 xanthomonas
37	166	8.4	550	10	Q7XU05	Q7XU05 oryza sativ
38	165	8.3	377	16	Q8PRD5	Q8PRD5 xanthomonas
39	163	8.2	364	2	Q9F0G8	Q9F0G8 rhizobium m
40	157.5	7.9	363	2	Q07652	Q07652 cellvibrio
41	155.5	7.8	516	2	Q66185	Q66185 bacillus ci
42	154.5	7.8	722	16	Q97L56	Q97L56 clostridium
43	153	7.7	357	2	Q59441	Q59441 fibrobacter
44	151.5	7.6	356	16	Q9PAL2	Q9PAL2 xylella fas
45	149.5	7.5	329	16	Q9X274	Q9X274 thermotoga

ALIGNMENTS

RESULT 1

Q8RP23 PRELIMINARY; PRT; 397 AA.

ID Q8RP23; AC Q8RP23; DT 01-JUN-2002 (TREMELrel. 21, Created)

DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE Endo-beta-1,4-glucanase

OS Paenibacillus sp. KCTC8848P.

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.

OX NCBI\_TaxID=109199;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KCTC 8848P.

RA Park J.N., Kim H.O., Shin D.J., Lee H.B., Chun S.B., Bai S.;

RT Cloning of Paenibacillus sp. endo-beta-1,4-glucanase gene and its co-

RT expression with Endomyces fibuliger beta-glucosidase gene in

RT Saccharomyces cerevisiae."

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF345984; AAL83749.1; "

DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR001547; Glyco\_hydro\_5.

DR Pfam; PF00150; cellulase; I.

DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.

SQ SEQUENCE 397 AA; 44539 MW; 1736A4C0C04E298F CRC64;

Query Match 52.7%; Score 1044; DB 2; Length 397;

Best Local Similarity 54.7%; Pred. No. 1, 1e-74;

Matches 197; Conservative 44; Mismatches 103; Indels 16; Gaps 5;

QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHRSRDYRSMLEQIKSLGNTRL 63

Db 37 GYVHTQGNKIVDETGKEAFAFNGLNWFGLETPNYTLRLGLWSRMDMLDQVKEGYNLRL 96

QY 64 PYSDILKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSQ 123

Db 97 PYSNQLFDSSSRADSIDYIK-NPDLVJUTPTQIMDKLEKAGQGIQIILDRHRPDCSQ 155

**This Page Blank (uspto)**

DT	01-FEB-1991 (Rel. 17, Created)	QY	285	TTLOSTTDQW----	LKTLVOYLR-----	PTAQYGADSFQWTFWSNPDSGDTGGI	331
DT	01-FEB-1991 (Rel. 17, Last sequence update)	Db	392	TDLPDGEDQVSMSELSVSGEYVRSRILGEEYQIDRTPREEFSEVIWDFN--	DGTTQGF	449	
DE	28-FEB-2003 (Rel. 41, Last annotation update)	QY	332	LKDDWQ-----	TVDIVKD	344	
DE	Endoglucanase C precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).	Db	450	VQNSDPLDVTIENVND	466		
GN	CELUC.						
OS	Bacillus sp. (strain N-4 / JCM 9156).						
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.						
OX	NCBI_TaxID=1413;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE=89326127; PubMed=2666258;						
RA	Fukumori F., Kudo T., Sashihara N., Nagata Y., Ito K., Horikoshi K.;						
RT	"The third cellulase of alkalophilic Bacillus sp. strain N-4:						
RT	evolutionary relationships within the cel gene family.";						
RL	Gene 76:289-298(1989).						
CC	!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic						
CC	linkages in cellulose, lichenin and cereal beta-D-glucans.						
CC	!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl						
CC	hydrolases).						
CC	-----						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -						
CC	the European Bioinformatics Institute. There are no restrictions on its						
CC	use by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>						
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).						
CC	-----						
DR	EMBL; M25500; AAA22306.1; --						
DR	PIR; JS0174; JS0174.						
DR	HSSP; O85465; 3A3H.						
DR	InterPro; IPR005086; CBM_17_28.						
DR	InterPro; IPR008979; Gal_bind_like.						
DR	InterPro; IPR001547; Glyco_hydro_5.						
DR	Pfam; PF03424; CBM_17_28; 2.						
DR	Pfam; PF00150; cellulase; 1.						
DR	PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.						
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.						
FT	SIGNAL						
FT	CHAIN						
FT	ACT_SITE						
FT	ACT_SITE						
SQ	SEQUENCE						
	Query Match						
	Best Local Similarity						
	Matches						
QY	13 ILDANNVPVRIA-----GINWFGFETCNVYVHGHSRSDYRSMLDQIKSLGYNTIRLP-YS	66					
Db	87 LADQDGVPIQLRGHSTHGLQWFG-EIVNENAFALANDW-----GSNVIRLALYI	135					
QY	67 DDILKPGTMPSINFYQMNQDLOGLTSLQVNDKIVAYAG-----QIGLRILDRHDPDCS	121					
Db	136 GE-----NAYRNPL-----IEKV--YAGIELAKENDMYIIDWHVHAPG	174					
QY	122 GQSLAWYTSVSEATWISDLOA-----LAQRYKGNPTVVGFDLHNEPHDPACWGC	171					
Db	175 DPNADIYQGGVNE-----DGEYILGAKDFFLHIAEKYPNDPHLI-YELANEPSSNSGGP	228					
QY	172 G-----DPSIDWRLAERAGNA-----VLSVNP	195					
Db	229 GITNDEGWEAREYAPIVD---ALRDSGNAEDNIIIVGSPNWSQRMDLAADNPIDDH	285					
QY	196 ----LLIFVEGVQSYNGSYWVG-----GNLQAGQVPPVNVNPNRLVYSAHDYATSVY	245					
Db	286 HTMYTLHFYTGTHGTSYEPGIGSSDRSNVMAKAY-----ALDKGKAIF	332					
QY	246 PQTWESDFTFNNMP-----GIW-----NK-----NWGLFNON-----IAPVWLGEF-G	284					
Db	333 ATEWGVSEADGNGNPYNLEADVWLNFLNENNISWTNWS-LTNKNETSGATPFFILNESDA	391					

Search completed: August 2, 2004, 16:34:56  
Job time : 9.33333 secs





Query Match 7.4%; Score 147.5; DB 1; Length 814;  
 Best Local Similarity 20.7%; Pred. No. 0.00061;  
 Matches 80; Conservative 61; Mismatches 134; Indels 111; Gaps 19;

QY 6 WHTSGRIIDANNVPRIAG-----INWFGFETCNVYVHGL--RSRDYRSM 49  
 DB 32 FQVSGTKLLDASGNELVWRGMRDISAIDLKKEIKIGNLGNLTLDAPTETAWGNPRTTKAM 91  
 QY 50 LDOIKSLGYNTIRLPYSDDILKPGTMPNSINFQMNQDLQGLTSLOWMDKIVAYAGQI 109  
 DB 92 IEKREMGFNARVPVTDW-THIGPADP-----YKIDE-----AWLNRVEVNVYVLDGCM 141  
 QY 110 RIILDRHRPDCSGOSALWYTSSEATWISDLOA-----LAQRYKGNPTV 154  
 DB 142 YALINLHH-----DNTWIIPTVANEORSKEKLVKWEQATATRFKDYDDH 185  
 QY 155 VGFPLNHPHPACWCGDPSIDWRLAARAGNAVLNVNPNLLIFVEGVQSYNGDSYWMG 214  
 DB 186 LLETWNEPRE-----VGSP-MEWMGGTYENRDVINEFN--LAVVNTIRAS-----G 229  
 QY 215 GN-----LQAGQYFVVLN--VPN--RLVYSAHDYATSVYPT-----WFS- 252  
 DB 230 GNNDKRFILVPTNAATGLDVALNDLVIPNDSRVIVSIHAYSPYFAMDVNGFSYWGSDY 289  
 QY 253 --PTFPNNMGIWNNKNGYLPNQNIAFVWLGEFGTTLQSTDTQWLKTLV---QYLRT 306  
 DB 290 DKASLTSELDIYR-----FVKNGRAVITIGFG-----TIDKNLSRVAHAHYAREA 339  
 QY 307 AQYGADSFQMTFWSNPDSDGTGIL 332  
 DB 340 VSRGIAVFWMDNGYNEFGDAETVALL 365

RESULT 12  
 GUN1\_RUMAL STANDARD; PRT; 406 AA.  
 ID GUN1\_RUMAL  
 AC P16216;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase) (EG-I).  
 GN EG I.  
 OS Ruminococcus albus.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;  
 OC Ruminococcus  
 OX NCBI\_TaxID=1264;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-58.  
 RC STRAIN=F-40;  
 RX MEDLINE=90078126; PubMed=2687251;  
 RA Ohniva K., Kajino T., Kato A., Shimizu S.;  
 RT "Structure of a Ruminococcus albus endo-1,4-beta-glucanase gene."  
 RL J. Bacteriol. 171:6771-6775 (1989)  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M30928; AAA26469.1; -  
 CC PIR; A43722; A43722.  
 CC HSPG; P17901; IEDG.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC Pfam; PF00150; cellulase; 1.  
 CC PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
 FT SIGNAL 1 43  
 FT CHAIN 44 406 ENDOGLUCANASE I.  
 FT ACT SITE 210 210 PROTON DONOR (BY SIMILARITY).  
 FT ACT SITE 330 330 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 406 AA; 45390 MW; 2E0172437B14FEA8 CRC64;

Query Match 7.3%; Score 145.5; DB 1; Length 406;  
 Best Local Similarity 22.9%; Pred. No. 0.00037;  
 Matches 88; Conservative 49; Mismatches 152; Indels 95; Gaps 18;

QY 15 DANNVPVRIAGIN-----WFGFETCNVYVHGLRSR-----DYSR 47  
 DB 49 ETENVVPSQTHNDTMTVTSAKDLVAKMTNGWNLGNTMDATAQGLGSEVSMLPLKVTNK 108  
 QY 48 SMLDOIKSLGYNTIRLPYSDDILKPGTMPNSINFQMNQDLQGLTSLOWMDKIVAYAGQI 107  
 DB 109 YIMDLPEAGFNLRIPVS-----WGNHIIIDKYTSDPAMMDRVQ---EIVNYGIDN 157  
 QY 108 GLRIILDRHRPDCSGOSALWYTSSE-----ATWISDLOAQAQRYKGNPTVVGFD 158  
 DB 158 GLYVILNTHHEE-----WYMPKSEKDGIEEIKAVAQ---IADRFKGYDEHLIFE 206  
 QY 159 LHNEHPDPACWCGDPSIDWRLAARA-----GNAVLNVNPNLLIFVEGVOS---YNGDS 210  
 DB 207 GLNEP-----RLRGEAEWTGTSEAREIINEYKAFVETVRASGNGNDR 251  
 QY 211 YWGGNLOAGQYF--VVLNVP--NRLVYSAHDYAT--SVYPTQWFSPTFPNN--MPGI 262  
 DB 252 CLMITGYAASAYNNLSAIELEPDSKLIISVHAYLPYSFALDTKGTDPETAPEL 311  
 QY 263 NNKNWGYLFNQNIAPVWLGEFGTTLQSTDTQWLKTLVQYLRPTAQYAGDSFQMTFWSN 322  
 DB 312 F-EHLNELFISKGPVIVGFGTGNKENTEDR-VKLEDYLAATAAKYDIPCVWMDNYA-R 368  
 QY 323 PDSGDTGGILKD--DWQTVDTVK 343  
 DB 369 ICGNGENFLGNRADLEWYFPDLIE 392

RESULT 13  
 GUN1\_RUMAL STANDARD; PRT; 409 AA.  
 ID GUN1\_RUMAL  
 AC P23661;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase) (EGB).  
 GN CELB.  
 OS Ruminococcus albus.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;  
 OC Ruminococcus  
 OX NCBI\_TaxID=1264;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SY3;  
 RX MEDLINE=91066933; PubMed=2250649;  
 RA Poole D.M., Hazlewood G.P., Laurie J.I., Barker P.J., Gilbert H.J.;  
 RT "Nucleotide sequence of the Ruminococcus albus SY3 endoglucanase  
 RL genes celA and celB."  
 RL Mol. Gen. Genet. 223:217-223 (1990).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M30928; AAA26469.1; -  
 CC PIR; A43722; A43722.  
 CC HSPG; P17901; IEDG.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC Pfam; PF00150; cellulase; 1.  
 CC PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.

RN	[1]	SEQUENCE FROM N.A.
RC		STRAIN=AM;
RX		MEDLINE=92138626; PubMed=1735723;
RA		Huang J., Schell M.A.;
RT		"Role of the two-component leader sequence and mature amino acid sequences in extracellular export of endoglucanase EGL from Pseudomonas solanacearum.";
RT		J. Bacteriol. 174:1314-1323 (1992).
RL	[2]	
RN		SEQUENCE OF 1-112 FROM N.A., AND PARTIAL SEQUENCE.
RP		MEDLINE=89291722; PubMed=2738021;
RX		Huang J., Sukorodhanan M., Schell M.A.;
RT		"Excretion of the egl gene product of Pseudomonas solanacearum.";
RT		J. Bacteriol. 171:3767-3774 (1989).
RL	[3]	
RN		PROCESSING.
RP		MEDLINE=90307678; PubMed=2195024;
RX		Huang J., Schell M.A.;
RA		"Evidence that extracellular export of the endoglucanase encoded by egl of Pseudomonas solanacearum occurs by a two-step process involving a lipoprotein intermediate.";
RT		J. Biol. Chem. 265:11628-11632 (1990).
CC	-1-	CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.
CC	-I-	SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Probable).
CC	-I-	SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl hydrolases).
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC		-----
DR		EMBL; M84922; AAA61980.1; --
PUR		A42649;
DR		InterPro; IPR001547; Glyco_hydro_5.
DR		InterPro; IPR000437; Prok_LipoProt_S.
DR		Fam; PF00150; cellulase; 1.
DR		PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR		PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW		Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane; Lipoprotein; Signal; Palmitate.
FT	SIGNAL	1..19
FT	PROPEP	20..45
FT	CHAIN	46..426
FT	LIPID	20..20
FT	LIPID	20..20
FT	ACT_SITE	249..249
FT	ACT_SITE	361..361
FT	SEQUENCE	426 AA; 45578 MW; 51E13AD442CF4A8 CRC64;
QY		Query Match 8.4%; Score 167; DB 1; Length 426;
Dd		Best Local Similarity 23.9%; Pred. No. 8e-06;
Dd		Matches 81; Conservative 49; Mismatches 129; Indels 80; Gaps 18;
Qy	35	NVVVHGLRSRDYRSMLODKISLGNTLRPLPSDDILKPPTMPSINFVMNQ--DLQGLT 92
Dd	141	NIITPSADSVTY-----YKNKGMLVRLPFWRERLPQ-----TLNQVDANELS 184
Qy	93	SLOWMDKIVAYAGIGIRLIILDRHPDCSGOSALWY-----TSVSSEATWISDLQAQR 147
Dd	185	RLTGFNVAVTATGQ---TVLLDPH-----NYARYYGNVIGSSAVPNASVADFWRRLATQ 235
Qy	148	YKGNPTVGGFDLHNPHDPACWGCDPSIDRWLAERAGNAVLSTVNPLLIFVEGVSYN 207
Dd	236	FKSNPRVI-LGLMNEPMSM-----PTEQLSGANAEIAIRSANASNVPFVG-NAWT 286
Ov	208	GDSYWWGNGLOCAGQYPVV--LVNP-VNLVYSADHY-----ATSVPEQTWFSD 252

DOMAIN 281 748 CATALYTIC.  
 FT DISULFID 39 133 BY SIMILARITY.  
 FT ACT\_SITE 503 503 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 553 553 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 748 AA; 80197 MW; AC96104137932B76 CRC64;  
 Query Match 14.7%; Score 290.5; DB 1; Length 748;  
 Best Local Similarity 23.4%; Pred. No. 3.1e-15;  
 Matches 111; Conservative 63; Mismatches 136; Indels 165; Gaps 23;  
 QY 7 HTSGREILDANNVVRVIRAGINWFFETCN-----YVHGLRS- 43  
 Db 290 NTKGNLTQKGLLPARGC--NWFGLEGRHEFSDADNPAGMELVAGNWMVNNQSGG 347  
 QY 44 RDVRSMLDQIKSLGTYNIRLPYSDDIL---KPGTSPNSINFYQMNQDLOGLSLQVMDK 99  
 Db 348 RTIOQTWELKQOQITWLRPIAPOTLDANDPQGRSPNLKN---HQSIRCSNARQALD 403  
 QY 100 IVAVAGQIGRIILDRH-----RPDCSGQSALWY-----TSSVS 133  
 Db 404 FIKLADQNDIQIFDIHSCSYVWGRAGRLDAREPPYVDNRVGYDFTRREYSCSATNPS 463  
 QY 134 EAT-----WISDQLAAL--QRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAA 182  
 Db 464 SVTRIHYADKQKWLJANUREIAGLSAKLVSNLIGIDVFNPEYD-YTWA-----EHWGMV 516  
 QY 183 ERAGNAVLSVNPPLLI FVEGVQSYNGDSY-----WPGNLOQAGQY 223  
 Db 517 EEAQCAINVENPNMLIIVEGI--SANANTQGTPTDPTSVPEPHGSTDLPNPNWGNLYEAGAN 575  
 QY 224 PVLVNLP-NRLVYSAHDYATSVYPTWFSPT-----FENNMPG 261  
 Db 576 PP--NIPKDRLLFSPHTYGVSVFQVQFMDPAQTECAGLEGDRRAQARCRIVNPTVLEQ 633  
 QY 262 IWNNKNGVLFNQTAPVWMLGEGF-----TTIQSTTDQTLTKLVQYL 303  
 Db 634 GWEEHFGYRELGVG-LIGFEGGNWDPGAKSSQADRNASHITTVQDQWQQAASVYF 692  
 QY 304 RPTAQYAGDSFQWTFWNPDSGDTGILKDDWQTVDTVKGYLPIKKSIFPPV 358  
 Db 693 K---RKGINA---CYWSMNPESADTMG---W-----YLTTP-----WDEPV 722

## RESULT 9

GUN1\_RALSO STANDARD; PRT; 424 AA.  
 ID GUN1\_RALSO  
 AC P58599;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN EGL OR RSP0162 OR R50516.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RC MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Bottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choiseine N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Iavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Sigulier P., Thebaud P., Whalen R.M., Wincker P., Levy M.,  
 RA Weissbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

CC CC (Probable).  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AL646076; CAD17313.1; ALT\_INIT.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC InterPro; IPR000437; Prok\_lipoprot\_5.  
 CC Pfam; PF00150; cellulase; 1.  
 CC PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 CC PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 CC Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;  
 KW Lipoprotein; Signal; Plasmid; Complete proteome; Palmitate.  
 FT SIGNAL 1 19 BY SIMILARITY.  
 FT PROPEP 20 43 BY SIMILARITY.  
 FT CHAIN 44 424 ENDOGLUCANASE.  
 FT LIPID 20 20 S-diacylglycerol cysteine (By  
 FT similarity).  
 FT LIPID 20 20 N-palmitoyl cysteine (By similarity).  
 FT ACT\_SITE 247 247 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 359 359 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 424 AA; 45080 MW; 5BF0D161AB7C87EF CRC64;  
 Query Match 9.4%; Score 186; DB 1; Length 424;  
 Best Local Similarity 25.5%; Pred. No. 2.5e-07;  
 Matches 86; Conservative 50; Mismatches 125; Indels 76; Gaps 19;  
 QY 35 NYVHGLRSRDYRSMLDQIKSLGTYNIRLPYSDDILKPGTSPNSINFYQMNQ--DLQSLT 92  
 Db 139 NYIYPSADSATY-----YKNGMNLVLPFRWERLP-----TLNQALDANELS 182  
 QY 93 SLOWMDKIVAYAGQIGRIILDRHDPDCSGQSALWY-----TSSVSEATWISDQLAQR 147  
 Db 183 RUTGFVNAVTAAG--TVLLDPH-----NYARYIGNVIGSSAVPNASVYFVPG-NAWT 233  
 QY 148 YKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAERAGNAVLSVNPPLLI FVEGVQSYN 207  
 Db 234 PKGNARVI-FGLMNEPSN-----PTEQWLSGANAALAAIRSANASNVVFPVPG-NAWT 284  
 QY 208 GDSYVWGNLOQAGQYFV--LVNPNR-LVYSAHDY-----ATSVYPTWFS--PT 254  
 Db 285 G-AHSWNCNQWYGTPTNGTYNKGINDPGRNLVFEVHQYLDGSDSGOSASCVSATIGAERLQD 343  
 QY 255 FENNMPGIGIWNKNWGYLFNQNIAPVWMLGEGFTTLOSTTDQTLTKLVQYLRPTAQYAGDSF 314  
 Db 344 FTN-----WLRNGY-----RGLGEGGAASNDTCQA-VANMLTFVKNADV----- 385  
 QY 315 QMTFWSWNPDSGDTGILK--DDWQTVDTVKGYLIAP 349  
 Db 386 -WTGMAWAGGFWPGWGYMYSIETSPNSGVKDPQVSVLAP 421  
 RESULT 10  
 GUN2\_RALSO STANDARD; PRT; 426 AA.  
 ID GUN2\_RALSO  
 AC P17974;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN EGL.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;

```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase B precursor (EC 3.2.1.4) (EGB) (Endo-1,4-beta-glucanase)
DE (Cellulase B).
GN CELB.
OS Clostridium thermocellum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OC Clostridium.
OX NCBI_TaxID=1515;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=NCIB 10682;
RC MEDLINE=86148508; PubMed=3453102;
RX Grepinet O., Beguin P.;
RA "Sequence of the cellulase gene of Clostridium thermocellum coding
RT for endoglucanase B.";
RL Nucleic Acids Res. 14:1791-1799 (1986).
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
CC GLUCANS.
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC hydrolases).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X03592; CAA27266.1; -.
CC FIR; A23512; C2CLBW.
CC DR HSSP; P54583; LECE.
CC DR InterPro; IPR002105; Dockerin_1.
CC DR InterPro; IPR002048; EF-hand_5.
CC DR InterPro; IPR001547; Glyco_hydro_5.
CC DR Pfam; PF00150; Cellulase; 1.
CC DR Pfam; PF00404; Dockerin_1; 2.
CC DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
CC DR PROSITE; PS00448; GLO_CELLULOSE_RPT; 2.
CC DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
CC FT SIGNAL 1 27
CC FT CHAIN 28 563 ENDOGLUCANASE B.
CC FT ACT_SITE 204 204 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 363 363 NUCLEOPHILE (BY SIMILARITY).
CC FT DOMAIN 502 557 2 X 24 AA APPROXIMATE REPEATS.
CC FT REPEAT 502 526 1.
CC FT REPEAT 534 557 2.
CC SQ SEQUENCE 563 AA; 63929 MW; 866FE55704ALDE4B CRC64;
Query Match 26.6%; Score 528; DB 1; Length 563;
Best Local Similarity 31.08; Pred. No. 4.5e-34;
Matches 130; Conservative 57; Mismatches 159; Indels 74; Gaps 13;
QY 1 AGGQY-----W-HTSGREILDANNVPVRIAGINWFGFETCNYYVHGLRSRDYSML 50
DB 27 AEGSYADLAPDDDLWHEGTNIVDKYGNKWITGANWFGFCNRCRMLDLSYHSDIADI 86
QY 51 DQIKSLGVNTRILPYSDDL---KPGTWMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQI 107
DB 87 ELVADKGINVVRMFIADLLYAMSQGIYPSTDTSYNNPALAGNSYELFNFMLENKRV 146
QY 108 GLRIILDRHPDCSQS---ALWYTSVSSEATWTSIDLOALAQRYKGNPTVVGFOLHNPH 164
DB 147 GIKVILDVHSPETNQGHNVPLWNTTITIEIFPKAWVAERYKNDTDTIGFDLKNPH 206
QY 165 -----DPACWCWCGDPSIDWRIAERAGNAVLVSNPNLLIFVEGVOSY----- 206

```

PFam; PF00404; Dockerin 1; 2.  
DR PROSITE; PS00018; EF HAND; UNKNOWN 2.  
DR PROSITE; PS00448; CLOS CELLULOSE\_RPT; 2.  
DR PROSITE; PS00659; GLYCOSYL HYDROL\_F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
FT SIGNAL 1 30  
FT CHAIN 31 566  
FT ACT SITE 226 226  
FT ACT SITE 226 226  
FT ACT SITE 381 381  
FT DOMAIN 503 549  
FT REPEAT 503 526  
FT REPEAT 536 549  
SQ SEQUENCE 566 AA; 63199 MW; 2CC9D9AD87C3178 CRC64;  
  
Query Match 30.1%; Score 596.5; DB 1; Length 566;  
Best Local Similarity 34.2%; Pred. No. 1.8e-39;  
Matches 139; Conservative 55; Mismatches 143; Indels 69; Gaps 15;  
  
QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNTIRLP 64  
Db 44 WLHCKGNKIYDMYGNVWLTCANWFGFNCSENCPHG-AMYDVKTILTSIADRGINILRIP 102  
QY 65 YSDIL-----KPG-----TMPNSINFYQMNQDLOG-----LTSLOWMDKIVAYAGOI 107  
Db 103 ISTEILYSWMIGKPNVSVSTASNPYHVVPDYPDPETDDVKNSEIFDIINGYCKEL 162  
QY 108 GLRIILDRHRPDC--SQGS-ALWY---TSS---VSEATWISDLQALAQRYKGNPTVVGF 158  
Db 163 GIKWIDHSDANNSHYELWYKSTCGVVTWKWIDTLVWLADKYKNDITIAFD 222  
QY 159 LHNPHDP-----ACWCGCDPSIDWRLAERAGNAVLNVNPLLI FVEGQSY- 206  
Db 223 LKNEPHGKRGYTABVPKLAKWKNSTDENNKVAAETCAKALLEVNPKNVLIIEGVEQYP 282  
QY 207 -----NGD-----SYWNGNLOGAGQYVVLNPN-RLVYSADHYATSV 244  
Db 283 KTEKGYTIDPDIWATGADSPWASWNGNLRGVKDYPIDLGPLNSQIVSVPHDYGPSV 342  
QY 245 YPQTWF-SDPTFFNMPGIVNNKGYLFNQIAPVWLGEFTLQSTDTDTWTLVQYL 303  
Db 343 YAPWFEKDFMTQTLDDYDTWAYTHDQIAPILIGEWGMDGKGNKMTLLRDYI 402  
QY 304 RPTAQYADSPQWTFWSPNPDGDTGILKDDWQVTDVTDGYLAP 349  
Db 403 -----VQNRHHTFWCINPNSGDTGILLGNDSWTDEAKVALLKP 442  
  
RESULT 6  
ID\_GUNB\_CELFI STANDARD; PRT; 747 AA.  
AC P50400;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DB (Cellulase).  
GN CEND.  
OS Cellulomonas fimi.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Micrococineae; Cellulomonadaceae; Cellulomonas.  
OX NCBI\_TaxID=1708;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93209933; PubMed=9458833;  
RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.;  
RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase  
D (CEND), a family A beta-1,4-glucanase."  
RL J. Bacteriol. 175:1910-1918 (1993).  
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -1- PATHWAY: Cellulose degradation.  
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.  
CC -1- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)

domain.  
CC -1- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
hydrolases).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; L02544; AAA23089.1; -.  
DR HSSP; P07986; 1EXG.  
DR InterPro; IPR001919; Bac\_celose-bind.  
DR InterPro; IPR008965; Cellul bind.  
DR InterPro; IPR008957; FN-III-like.  
DR InterPro; IPR003961; FN-III.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00553; CBM\_2; 1.  
DR Pfam; PF00150; cellulase; 1.  
DR Pfam; PF00041; fn3; 2.  
DR SMART; SM00637; CBD II; 1.  
DR SMART; SM00600; FN3; 2.  
DR PROSITE; PS00659; GLYCOSYL HYDROL\_F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
FT SIGNAL 1 39  
FT CHAIN 40 747  
FT DOMAIN 448 542  
FT DOMAIN 546 639  
FT ACT SITE 208 208  
FT ACT SITE 349 349  
FT ACT SITE 349 349  
SQ SEQUENCE 747 AA; 78936 MW; BD15473CD8B42BD CRC64;  
  
Query Match 27.3%; Score 541.5; DB 1; Length 747;  
Best Local Similarity 34.7%; Pred. No. 5.5e-35;  
Matches 131; Conservative 59; Mismatches 144; Indels 45; Gaps 15;  
  
QY 1 AGGGYV-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYN 59  
Db 40 ATGDDMLHVEGNTIYDSTGKAILSGVNVWFGFNASERVHGLWSGNITQITQQWAQRGIN 99  
QY 60 TIRLEYSDDIL---KPGTW---PNSINFYQMNQDLOGLTSLQWMDKIVAYAGQIGRLILD 114  
Db 100 VVRVPVSTQLLEWKAGTFLKEN-VNTY-ANPELGKNSLQIFEYWLTLCKQYGIKVELD 157  
QY 115 RH--RPDCSGQ-SALWYTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPH-----D 165  
Db 158 VHSAEADNSGHVYNNWVWKGDIITEDVYEGVWAAATRWKDDITVIGADIKNPEPHGTQGST 217  
QY 166 PACWCGDPSIDWRLAERAGNAVLNVNPLLI FVEGQSYN-----GDSY--W 212  
Db 218 RAKNDGTTDKONFKFAETASKKILAINPNWLVFVEGVLEIYKGVPTWSTGLTDYGTW 277  
QY 213 WGNLQAGQVVPVLNV-PNRLVYSAHYATSVPTWTF-SDPTFFNMPGIVNNKGYL 270  
Db 278 WGNLGRVDRDHPIDLGAHQDQVYSPHDYGLVFDQKWFQKDFDKASLTADWVGNWLF 337  
QY 271 FNQNIAPVWLGEFGTTL-QSTTDQW---LKTVLQYLSPTAQYGDADSFQWTFWSPNPDG 326  
Db 338 HDEDIAPLLIGWGRGLQDPRQDKWMAALRDLDVAERR-----LSQTFVWLNPNSG 388  
QY 327 DTGGILKDDWQVTDVTK 343  
Db 389 DTGGLLDWKTWEVK 405  
  
RESULT 7  
GUNB\_CLOTH STANDARD; PRT; 563 AA.  
ID\_GUNB\_CLOTH  
AC P04956;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)  
Endoglucanase/exoglucanase B precursor [includes: Endoglucanase  
(EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase)  
(Cellulohydrolase); Exoglucanase (EC 3.2.1.91) (Exocellobiohydrolase)  
(1,4-beta-cellobiohydrolase)].

Caldoceum saccharolyticum (Caldicellulosiruptor saccharolyticus).  
Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;  
Caldicellulosiruptor.  
NCBI\_TaxID=44001;  
[1]

SEQUENCE FROM N.A.  
MEDLINE=89098398; PubMed=2789517;  
Saul D.J., Williams L.C., Love D.R., Chamley I.W., Bergquist P.I.;  
"Nucleotide sequence of a gene from Caldocellum saccharolyticum  
encoding for exocellulase and endocellulase activity";  
Nucleic Acids Res 17:439-439 (1989).

-!- FUNCTION: THIS PROTEIN IS MADE UP OF TWO DOMAINS: THE N-TERMINAL  
DOMAIN HAS EXOGLUCANASE ACTIVITY WHILE THE C-TERMINAL DOMAIN IS  
AN ENDOGLUCANASE.

-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.

-!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages  
in cellulose and cellotetraose, releasing cellobiose from the non-  
reducing ends of the chains.

-!- SIMILARITY: In the N-terminal section; belongs to cellulase family  
F (family 10 of glycosyl hydrolases).

-!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CELLULASE FAMILY  
A (FAMILY 5 OF GLYCOSYL HYDROLASES).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----  
EMBL; X13602; CAA1936.1; --  
PIR; S02711; S02711.  
HSP; Q06851; INBC.  
InterPro; IPR001956; CBD 3.  
InterPro; IPR008965; Cellul bind.  
InterPro; IPR001000; Glyco\_hydro\_10.  
InterPro; IPR001547; Glyco\_hydro\_5.  
Pfam; PF00942; CBW\_3; 1.  
ProDom; PD001947; CBD 3; 1.  
SMART; SM00633; Glyco\_10; 1.  
PROSITE; PS00591; GLYCOSYL\_HYDROL\_F10; 1.  
PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
Cellulose degradation; Hydrolase; Glycosidase; Repeat;  
Multifunctional enzyme; Signal.  
SIGNAL 1 28  
CHAIN 29 1039  
ENDOGLUCANASE/EXOGLUCANASE B.  
THR/PRO-RICH, TANDEM REPEATS OF T-P.  
CELLULOSE-BINDING (BY SIMILARITY).  
DOMAIN 376 416  
DOMAIN 417 570  
THR/PRO-RICH, TANDEM REPEATS OF T-P.  
DOMAIN 571 618  
PROTON DONOR (POTENTIAL).  
ACT\_SITE 177 177  
NUCLEOPHILE (BY SIMILARITY).  
ACT\_SITE 285 285  
BY SIMILARITY.  
ACT\_SITE 792 792  
SEQUENCE 1039 AA; 117641 MW; OE0378171594DDAE CRC64;  
Query Match 31.38; Score 620.5; DB 1; Length 1039;  
Best Local Similarity 35.58; Pred. No. 5.1e-41;  
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFTCNVYVHGRSDYRSMLDQIKSLGNTIRLP 64  
DB 629 WLYVSGNKIVKDGPRPWLGTGINWFGYNTGNVFDGWSCNLTKLAETIANRGNLLRVP 698

QY 65 YSDIL---KPGTVPN-SINFYQNMODLOGLTSLQWMDKIVAVAGOIGRIILDRH--RP 118  
DB 689 ISAEILNLNQQGIYKPKNINYY-VNPELEGKNSLEVFVDIVVQTCKEVGLKIMLDIHSIKT 747  
QY 119 DCSQ-SALAWYTSVSSEATWISDLOLAQRYKGNPTVVGPDILHNEPDP-----ACWG 170  
DB 748 DAMGHIVPVWIDEKFTPEDFYKACEWITNRYKDDTIIIAFDLKNRPHGKQWDTTAKWD 807  
QY 171 CGPDSIDWRLAAERAGNAVLNVNPNLIFVGVQSYNGD-----SYWGGNL 217  
DB 808 NSTDINNWKYAAETCAKRIILNINFLNLLIVIEGIEAYPKDDVTWTSKSSSDYYSWTWGGNL 867  
QY 218 QGAGCPVVL-NVFNRLVYSAHDYATSVYPTWESDPTFPNN--MPGIWKNKGYLPNQ 274  
DB 868 RGVKXYPINIGKYQNVVPHDYGFSVYQOPWFY-PGFTKESLLQDCWPNWAYIMEEN 926  
QY 275 IAPVWLGEFGTTLQSTTDQTLKTLVQYLRTTAQYCADSFQWTFWSNPDGDTGILKD 334  
DB 927 IAPLLIGWGGHLDGADNEKWKYLRDYII-----ENH1HHTFWCFNANSNGDTGGLVGY 980  
QY 335 DWQTVDTVKGYLAP 349  
DB 981 DFTWDEKKYSFLKP 995

RESULT 5  
GUNG\_CLOTH STANDARD; PRT; 566 AA.  
AC Q05332;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase G precursor (EC 3.2.1.4) (Egg) (Endo-1,4-beta-glucanase)  
DE (Cellulase G).  
GN CELG.  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1515;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCIB 10682;  
RX MEDLINE=93273701; PubMed=8501039;  
RA Lemaire M., Beguin P.;  
RT "Nucleotide sequence of the celG gene of Clostridium thermocellum and  
RT characterization of its product, endoglucanase CelG";  
RL J. Bacteriol. 175:3353-3360(1993).  
CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-  
CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
CC GLUCANS.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS  
CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN  
CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
CC hydrolases).  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

-----  
EMBL; X69390; CAA49187.1; --  
PIR; A40589; A40589.  
HSP; P54583; LECE.  
InterPro; IPR002105; Dockerin\_1.  
InterPro; IPR002048; EF-hand.  
InterPro; IPR001547; Glyco\_hydro\_5.  
Pfam; PF00150; cellulase; 1.







PRO/SER/THR-RICH (LINKER).  
CELLULOSE-BINDING (BY SIMILARITY).  
PROTON DONOR.  
NUCLEOPHILE.

401 461  
402 562  
203 203  
323 323  
75 161  
209 212  
48 50  
51 52  
53 55  
57 58  
61 62  
65 68  
72 72  
73 73  
75 76  
80 84  
87 96  
97 98  
101 107  
108 111  
113 114  
127 129  
130 131  
134 147  
148 149  
151 158  
160 160  
161 162  
163 163  
175 188  
189 191  
193 194  
195 199  
206 207  
209 209  
215 217  
218 218  
219 230  
231 233  
235 236  
238 242  
244 244  
246 247  
248 249  
250 251  
254 255  
257 257  
259 263  
274 279  
283 285  
289 292  
294 299  
300 307  
308 308  
309 313  
314 315  
319 323  
323 323  
331 343  
344 344  
347 350  
351 352  
356 359  
367 369  
373 373  
375 376  
380 380  
382 387  
388 388  
389 391  
562.AA; 60747 MW; 84E6256406A35041 CRC64;

Query Match 99.6%; Score 1974; DB 1; Length 562;  
Best Local Similarity 99.7%; Pred. No. 7.7e-148;

Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVVRVRIAGNWFGEFCTNYYVHGLSRDYRSMLDQIKSLGYNT 60  
DB 42 AGGYWHTSGREILDANNVVRVRIAGNWFGEFCTNYYVHGLSRDYRSMLDQIKSLGYNT 101  
QY 61 IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGRIILDRHRPDC 120  
DB 102 IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGRIILDRHRPDC 161  
QY 121 SGQSALWYTSVSSEATWISDLQALQRYKGNFTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 162 SGQSALWYTSVSSEATWISDLQALQRYKGNFTVVGFDLHNEPHDPACWCGDPSIDWRL 221  
QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGNGNLQAGQYFVVLNVNRLVYSAHDY 240  
DB 222 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGNGNLQAGQYFVVLNVNRLVYSAHDY 281  
QY 241 ATSVPQTFWSDPTFPNNMFGIWNKNGWYLFNQNIAPVWLGEFGTTLQSTTDQTLWKTIV 300  
DB 282 ATSVPQTFWSDPTFPNNMFGIWNKNGWYLFNQNIAPVWLGEFGTTLQSTTDQTLWKTIV 341  
QY 301 QYLRPTAQYGADSFQWTFWSNPDSDGTGGILKDDWQTVTDVTKDGYLAPIKSSIFDPV 358  
DB 342 QYLRPTAQYGADSFQWTFWSNPDSDGTGGILKDDWQTVTDVTKDGYLAPIKSSIFDPV 399

RESULT 2.

GUN\_PABPO STANDARD; PRT; 397 AA.

ID AC P23548; 01-NOV-1991 (Rel. 20, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).  
OS Paenibacillus polymyxa (Bacillus polymyxa).  
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.  
OX NCBI\_TaxID=1406;  
RN [1]  
RP MEDLINE=90170877; PubMed=2307659;  
RX Baird S.D., Johnson D.A., Seligy V.L.;  
RA "Molecular cloning, expression, and characterization of  
RT endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus  
RT circulans";  
RL J. Bacteriol. 172:1576-1586(1990).  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
CC hydrolases).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M33791; AAA22631.1; -.  
DR F1R; A35136; A35136.  
DR HSSP; P54583; LECE.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00150; cellulase; 1.  
DR PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.  
KW Cellulose degradation; Hydrolase; Glycosidase.  
FT ACT\_SITE 194 194 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 317 317 NUCLEOPHILE (BY SIMILARITY).  
SQ SEQUENCE 397 AA; 44357 MW; B9C2B80C04F0A2A CRC64;

Query Match 52.3%; Score 1036; DB 1; Length 397;  
Best Local Similarity 54.2%; Pred. No. 3.1e-74;  
Matches 193; Conservative 43; Mismatches 108; Indels 12; Gaps 4;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:59 ; Search time 8.33333 Seconds  
(without alignments)  
2236.934 Million cell updates/sec

Title: US-09-997-504A-12  
Perfect score: 1982  
Sequence: 1 AGGGYWHTSGREILDANNVP.....VDTVDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1974	99.6	562	1	GUN1_ACICE
2	1036	52.3	397	1	GUN_PAEPO
3	792.5	40.0	484	1	GUN_A_XANCP
4	620.5	31.3	1039	1	GUNB_CALSA
5	596.5	30.1	566	1	GUNB_CLOSTM
6	541.5	27.3	747	1	GUNB_CELFI
7	528	26.6	563	1	GUNB_CLOTM
8	290.5	14.7	748	1	GUNB_PSEFL
9	186	9.4	424	1	GUN1_RALSO
10	167	8.4	426	1	GUN2_RALSO
11	147.5	7.4	814	1	GUNB_CLOSTM
12	145.5	7.3	406	1	GUN1_RUMAL
13	134	6.8	409	1	GUNB_RUMAL
14	134	6.8	584	1	GUNB_CLOCE
15	132	6.7	825	1	GUN3_BACS4
16	130.5	6.6	900	1	GUNH_CLOSTM
17	128	6.5	448	1	GUN_CLOSA
18	123	6.2	336	1	GUNB_RUMFL
19	123	6.2	440	1	GUNB_CLOCL
20	121.5	6.1	658	1	GUN3_PIBSU
21	120	6.1	517	1	GUN3_CLOLO
22	119	6.0	438	1	EXG_CANAL
23	117	5.9	501	1	YBQ6_YEAST
24	115	5.8	388	1	GUN3_HUMIN
25	115	5.8	562	1	EXG2_YEAST
26	112.5	5.7	418	1	GUN2_TRIRE
27	111.5	5.6	466	1	GUN5_THSEFU
28	111.5	5.6	532	1	HEXA_DICDI
29	111	5.6	764	1	YIS7_YEAST
30	109.5	5.5	475	1	GUNB_CLOCE
31	108.5	5.5	1024	1	SVIP_STAAU
32	107	5.4	312	1	GUN4_RUMAL
33	107	5.4	515	1	GUNB_CLOCL

34	106.5	5.4	505	1	GUNV_ERWCA
35	106.5	5.4	547	1	GUN1_BUTFI
36	106.5	5.4	800	1	GUN_BACSI
37	105.5	5.3	364	1	GUNB_RUMAL
38	105.5	5.3	504	1	GUNV_ERWCA
39	104.5	5.3	429	1	GUNB_BUTFI
40	104.5	5.3	459	1	GUNA_STRLI
41	103.5	5.2	445	1	SPRI_YEAST
42	103	5.2	620	1	TYRO_NEUCR
43	102.5	5.2	522	1	MAL2_DROME
44	102	5.1	444	1	GUNV_ERWCA
45	99	5.0	383	1	MANA_STRLI

ALIGNMENTS

RESULT 1  
GUN1\_ACICE  
ID GUN1\_ACICE STANDARD; PRT; 562 AA.  
AC P54583; 1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)  
DE (Cellulase E1) (Endocellulase E1).  
OS Acidothermus cellulolyticus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Frankineae; Acidothermaceae; Acidothermus.  
OX NCBI\_TaxID=28049;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 43068 / 11B;  
RA Laymon R.A., Himmel M.E., Thomas S.R.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.  
RX MEDLINE=96346058; PubMed=8718854;  
RA Sakon J., Adney W.S., Himmel M.E., Thomas S.R., Kaplun P.A.;  
RT "Crystal structure of thermostable family 5 endocellulase E1 from  
Acidothermus cellulolyticus in complex with cellobiose";  
Biochemistry 35:10648-10660(1996).  
CC -!- FUNCTION: Has a very high specific activity on  
carboxymethylcellulose.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- MISCELLANEOUS: Thermostable enzyme with an optimal temperature of  
81 degrees Celsius.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
hydrolases).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
EMBL; U33212; AAA75477.1; ;  
DR PDB; 1E0E; 14-OCT-96.  
DR PDB; 1C0D; 23-JUL-99.  
DR InterPro; IPR001919; Bac celose-bind.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00553; CBM\_2; 1.  
DR Pfam; PF00150; cellulase; 1.  
DR SMART; SM00637; CBD II; 1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
DR Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.  
FT SIGNAL 1 41  
FT CHAIN 42 562  
FT DOMAIN 42 400  
FT CATALYTIC.

This Page Blank (uspto)

wed Aug 4 10:02:19 2004

us-09-997-504a-12.rpr

```
QY 68 DILKPGTTPNSINFYQ-----MNQDLQG-----LTSIQVMDKIVAVAGQIGLRI 111
Db 66 DLTNETLANNVTVROSFGSLGNDIVGQTNPNPSIIDPLIEAYKTVTTLGNNDVMV 125
QY 112 ILDRH--RPD--CSGQ--SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD 165
Db 126 ILNHLTKPCWCANDGNGFFGQDFDPTVVAALKQMAATFNGSVNVGMSLRNELRG 185
QY 166 PACWCGDPSIDWRLAARAGNAVLSNPNLLIFVEGVQSYNGDSYWGNGNLQAGQYPV 225
Db 186 PK-----QNVNDWFKYMQQGAEAHSAANNKVLVLSGL-SFDAD-----LSFVRSRPV 232
QY 226 VLNVNRLVYSAHYATSVYPQTWFS-----PTFPNNMPG-IWNK---NWGYLFNQNTA 276
Db 233 KLSFTGKLVFELHWS-----FSDGNSWAANNPNNDICGRVLRNIGNGGGYLLNQGF- 283
QY 277 PVMLGEGFTTLQ--STDTQWLKTLVOYLPTAQYAGDSFQWTFW-----SWNPDSGDTG- 329
Db 284 PLFLSEFGIDERGVTNDNRYFGCL-----TGMAAENDVDWSLWALTGSYLLRQGVKM 337
QY 330 ----GILKDDMQTV 339
Db 338 NEYGVLDSDWISV 351
```

```
QY 330 -GILKDDMQTV 339
Db 365 YGMLDANWHNV 375
```

Search completed: August 2, 2004, 16:37:39  
Job time : 13.3333 secs

```
RESULT 15
T51476
hypothetical protein K3M16_70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51476
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew
submitted to the Protein Sequence Database, August 2000
A:Reference number: 225394
A:Accession: T51476
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-526 <SAT>
A:Cross-references: EMBL:AL391150
A:Experimental source: cultivar Columbia; BAC clone K3M16
C:Genetics:
A:Map position: 5
A:Introns: 135/3; 218/2; 327/2; 393/1
A>Note: K3M16_70
```

```
Query Match 10.8%; Score 214; DB 2; Length 526;
Best Local Similarity 23.2%; Pred. No. 5.6e-09;
Matches 86; Conservative 60; Mismatches 161; Indels 64; Gaps 14;
QY 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLDQIKSLGYNIRLPY-- 65
Db 30 TKSRWIVNNKGRHVKACANW-PSHLKPVVAEGLSQPMDSISKIKDMGFNCVRLTWPL 88
QY 66 ----SDDLKPGTWPNSINFQNMQDLQGLTS-----LQVMDKIVAVAGQIGLRI 111
Db 89 ELMINDTLAFNVTYKQSFERYGLDHELGQIVTHNPYIWTPLINVFQAVVYSLGRHDVMV 148
QY 112 ILDRHP-----DCSGQ--SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD 165
Db 149 ILDNHKTVPFGCCSNDDPDPAFFGDPKFPNLDLMLGLKQMATIFMNVKRVVVGMSLRNELR- 207
QY 166 PACWCGDPSIDWRLAARAGNAVLSNPNLLIFVEGVQSYNGDSYWGNGNLQAGQYPV 225
Db 208 ----GYNHTSKDWKYMKQGAEAHVTNPNVLVILSGIN-----FDADLSFLKDRPV 255
QY 226 VLNVNRLVYSAHYATSVYPQTWFSPT--FPNNMFGIWNKXWGYLFNQNTAPVWLGEF 283
Db 256 NLSFKKLVLELHWYSFTDGTGQWKNHNVNDFCSQMFSEKRRRTGTFVLDQGF-PLFLSEF 314
QY 284 GT-----TLQSTTDQTLKTLVOYLPTAQYAGD-SFQWTFWS-----WNPDSGDTG---- 329
Db 315 GTDQGGDLEGN-----RPNMCLAWAEKDLDAWYAVTGYYFREGKGVVEA 364
```

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. sub-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000

A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins E.A.;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menco, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaske A.;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira M.; Tshakoto, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A;Reference number: A59328

A;Contents: annotation

C;Genetics:

A;Gene: XF0810

Query Match 16.5%; Score 328; DB 2; Length 565;  
Best Local Similarity 29.6%; Pred. No. 7.4e-18;  
Matches 103; Conservative 54; Mismatches 135; Indels 56; Gaps 12;

Qy	12	EILDANNV-----PVRIAGIKNWFPETCNYYVHGLRSRDYRSMLDQIKSLGYNTRL	63
Dd	4	ETMDAQNAVSDTHFVEPELH--GVNWRGLETAQHLLQLGDQRWRFVLDMQSILGINAIRL	61
Qy	64	PYSDDILKEGTMPNSINIFYOMKNOLOGTSLSQVMDKIVAVAGOIGURITILDRHPDC-SG	122
Dd	62	PLCSDTLH-GTMPNLNDLVR-NPDLKGRTALQIADAIIDEAGKGRNRVLLAHGYECPTD	119
Qy	123	QSALWYTSSVSSEATWISDLQALAORYKGNPW-VGFDLHNPHDPACWGCGPSIDRWLA	181
Dd	120	GNPILLRSVDSESHOWISDVQFITSHYRAQQKVNVGVLDAMNHRPFQSGSGSTPDPNRV	179
Qy	182	AERAGNAVLNVNPLLIIFVEGV----QSYNGDSYWVGNGNLQGAGQYFVVLYPNR-LVYS	236
Dd	180	VERAAAILAMTPWLIGIQVPVLGNPPCLDSAPISDDNIQ---SQHCVPILPARNLLLM	237
Qy	237	AHDYATSVPQTWFSDFTPNNMFGIMWNKWGYLFNQNIAPVWLBFGF---TTLOSTTDQ	293
Dd	238	PRFAGTDTIDE-----AALGAFFSKOTVLPNSLDA	267

294 TWJ.KTILVOYLRPTAOYGADSFOWTFWSWNPDSGDTGGILKDDWOTVDT 341

D<sub>b</sub> 268 TDAEQLAHRIDALLAFGIR--OGFYGSWMTSAQMPFGMLDNDGRTPT 313

RESULT 13

S19552  
 cellobiohydrolase C - Pseudomonas fluorescens  
 C/Species: Pseudomonas fluorescens  
 C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 15-Oct-1999  
 C/Accession: S19552; S16849  
 R/Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.  
 Biochem. J. 279, 793-799, 1991  
 A>Title: The cellobiohydrolase from Pseudomonas fluorescens subsp. cellulosa consists of "A"  
 A/Reference number: S19552; MUID:92061996; PMID:1953673  
 A/Accession: S19552  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A/Residues: 1-748 <FER>  
 A/Cross-references: EMBL:X61299  
 R/Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.  
 submitted to the EMBL Data Library, August 1991  
 A>Description: The cellobiohydrolase from pseudomonas fluorescens subsp cellulosa consists  
 A/Reference number: S16849  
 A/Accession: S16849  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A/Residues: 1-262, 'C', 264-291, 'K', 293-748 <FE2>  
 A/Cross-references: EMBL:X61299; MID:g45501; PID:g45502  
 C/Genetics:  
 A/Gene: celC  
 C/Superfamily: bacterial cellulose-binding domain homology; glycosidase GWGW domain hom  
 F/38-134/Domain: bacterial cellulose-binding domain homology <BC>

F;183-220/Domain: glycosidase GWGW domain homology <GWG>  
P;39-133/Disulfide bonds: #status predicted

Query Match	14.7%	Score	290.5	DB 2	Length	748
Best Local Similarity	23.4%	Pred. No.	9.3e-15			
Matches	111	Conservative	63	Mismatches	136	Indels
Gaps	23					
QY	7	HTSGREILDANNVPVRIAGINWFGFTCN	-----YVVHGLRS- 43			
Db	290	NTCGNLTKDQQLPARCG-NWFGLEGRHPSNDADNPSCGAPMELYAGNWMWVNNSSQGS	347			
QY	44	RDYRSMLDQIKSGYNTIRLPYSDDIL-----KGTWPNNSINFQMKQDLOGLTSLQWMDK	99			
Db	348	RTIQCTMTLKKQIGITMLRFLPAPQTLDANDPQGRSPNLKN-----HQSTQSNARQALED	403			
QY	100	IVAYAGIGILRIILDRH-----RPDCSGQSALWY-----TSSVS	133			
Db	404	FIKLADQNDIQIFIDHSCSYVWGRAGRLDARPPYVDANRVGYDFTREEYSCSATNNPS	463			
QY	134	EAT-----WISDQALAA--ORYKGNPTVVGFDLHNEPDPACWCGGDPSIDWELAA	182			
Db	464	SVTRIHAYDKQKWLANREIAGLSAKLGVSNLIGIDVFNEPYD-YTWA-----EMKGMV	516			
QY	183	ERAGNAVLSVNPMLLIFFVEGVQSYNGDSY-----WGGNLOQAQGY	223			
Db	517	EEAYQAINFVNPMLLIIVEGI-SANANTQGTPTDTSVPVPHGSTDLNPNWGENLYEAGAN	575			
QY	224	PVVLNVLP-NRLVYSAHDYATSVYPQTWFSDDT-----FPNNMGP	261			
Db	576	PP-NTPKDSLLFSPHTYGPSVFQRFQMDPAQTECAGLEGDEAAQARCRIVINPTVLEQ	633			
QY	262	INNKWGYLFENQIAPVWLGEFG-----TTLQSTTDQTLWLTLYQYL	303			
Db	634	GWEHFGLYLBELGVG-ILIFEGCNMDWPGAKSQSQADRNASHTTTINVDOQWQQAASYSF	692			
QY	304	RPTAQYGADSFQMTFWSNWPNDSGDTGGILKDDMQTVTDVKDGYLAPIKSSIFDVP	358			
Db	693	K--RKGINA-CYKWSMNPESADTWG-----W-----YLTG-----WDVP	722			

RESULT 14

E86265

Protein F3P19.15 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C/Accession: E86265

R/Theologian, A.; Ecker, J.K.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.

R/Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; D

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.

C/Authors: Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.;

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: E86265

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-522 <STO>

A/Cross-references: GB:AE005172; NID:g4850396; PIDN:AAD31066.1; GSPDB:GN00141

C/Genetics:

A/Gene: F3P19.15

A/Map position: 1

	Query Match	12.7%	Score 252.5	DB 2	Length 522
Best Local Similarity	24.9%	Pred. No. 5.4e-12			
Matches 93	Conservative 59	Mismatches 151	Indels 71	Gaps 18	
Qy	8	TSGREILDANNVPVRIAGINWFFETCNVYVHGRLSRDYSRMLDQIKSLGNTNTRLRPSYD	67		
Db	7	TSSRWIVDENGRLVKLVCAW--PSHLQFVVAEGSLSKQPDVAKKIVEMGNCVRLTWPFL	65		

Db 343 YAFWFEKFTMQTLDDYWDYTWAYTHDQGIAPILLIGEWGGMGGKQKQKQWTLRLDYI 402  
 QY 304 RPTAQYAGDSFQWTFWSNWPSDGTGILKDDQWTVDTVDXGYLAP 349  
 Db 403 -----VQNRHHTFWCINPNSGDTGGLGNDWSTWDEAKYALLKP 442

## RESULT 10

B47093  
 Cellulase (EC 3.2.1.4) CenD - Cellulomonas fimi  
 N:Alternate names: endo-1,4-beta-glucanase  
 C:Species: Cellulomonas fimi  
 C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 06-Dec-1996  
 R:Meinke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.  
 J. Bacteriol. 175, 1910-1918, 1993  
 A>Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (CenD),  
 A:Reference number: B47093; MUID:93209933; PMID:8458833  
 A:Accession: B47093  
 A>Status: preliminary  
 A:Molecule type: DNA; protein  
 A:Residues: 1-747 <MBI>  
 A:Experimental source: ATCC 484  
 A>Note: sequence inconsistent with the nucleotide translation  
 A:Pathway: cellulose degradation  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
 C:Superfamily: bacterial cellulose-binding domain homology  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:644-745/Domain: bacterial cellulose-binding domain homology <BCB>  
 F:645-744/Disulfide bonds: #status predicted

Query Match 27.3%; Score 541.5; DB 2; Length 747;  
 Best Local Similarity 34.7%; Pred. No. 2.2e-34;  
 Matches 131; Conservative 59; Mismatches 142; Indels 45; Gaps 15;

QY 1 AGGGYW-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYRSMLDQIKSLGYN 59  
 Db 40 ATGDDMLVEGNTIVDTGKEALLSGVNVWFGFNASERVFGLWSGNTITQITQMAQRGIN 99  
 QY 60 TIRLPYSDIIL---KPGTM--PNSINFYQMNQDQGLTSQVMDKIVAYAGQIGLRIILD 114  
 Db 100 VVRYPVSTQLLEWAKGFTLKPEN-VNTY-ANPELEGKNSLIQIPEYMLTLCKQYGIKVFLD 157  
 QY 115 RH--RPPCSGO-SALWTSVSEATWSDLOALAQRYKNGNPTVVGFDLHNEPH-----D 165  
 Db 158 VHSAEADNSGHVYVMWKKGDITTEDVYEGWEAAATRWKDDTTIGADIKNEPHGTQGSTE 217  
 QY 166 PACWCGCDPSIDWRLAERAGNAVLSVNPILLIFVEGVQSYN-----GDSY--W 212  
 Db 218 RAKWDGTTDXNFKFAETASKKILANPNWLVFVSGVEIYPRGPVPTSTGLTDYGTW 277  
 QY 213 WGNLQAGQYPPVVLN-PNRLVYSADHYATSVYPTQWF--SDTPFNNMPEIWNKQWYL 270  
 Db 278 WGNLGRVGRDHPIDLGAFQDQLVYSPHDYGLVFDQKWFQDFDKASLTADVWGNPNWLF 337  
 QY 271 FNQNIAPFWLGEFGTTL-QSTTDQW---LKTLYOYLRLPTAQYAGDSFQWTFWSNWPSDG 326  
 Db 338 HDEDIAPLLIGEWGRGLQDPRQKQWALADLVAER-----LSQTFVWLVNPSG 388  
 QY 327 DTGGLKDDMQWTVTVK 343  
 Db 389 DTGGLLLDDWKTVDEVK 405

## RESULT 11

CZCUBM  
 cellulase (EC 3.2.1.4) B precursor - Clostridium thermocellum  
 N:Alternate names: endo-1,4-beta-glucanase B precursor  
 C:Species: Clostridium thermocellum  
 C>Date: 31-Dec-1999 #sequence\_revision 31-Dec-1989 #text\_change 18-Jun-1999

C:Accession: A23512  
 R:Grepinet, O.; Beguin, P.  
 Nucleic Acids Res. 14, 1791-1799, 1986  
 A>Title: Sequence of the cellulase gene of Clostridium thermocellum coding for endogluc  
 A:Reference number: A23512; MUID:86148508; PMID:3453102

A:Accession: A23512  
 A:Molecule type: DNA  
 A:Residues: 1-563 <GRE>  
 A:Cross-references: GB:X03592; NID:g40668; PIDN:CAA27266.1; PID:g40669

A:Comment: This secretory enzyme is part of a highly active and thermostable cellulase co  
 C:Genetics:  
 A:Gene: celB  
 A:Function:

A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans  
 A:Pathway: cellulose degradation  
 C:Superfamily: cellulase B; Clostridium cellulase repeat homology  
 C:Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharide de  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:26-563/Product: cellulase B #status predicted <MAT>  
 F:502-525/Domain: Clostridium cellulase repeat homology <CCR1>  
 F:534-557/Domain: Clostridium cellulase repeat homology <CCR2>

Query Match 26.6%; Score 528; DB 1; Length 563;  
 Best Local Similarity 31.0%; Pred. No. 1.7e-33;  
 Matches 130; Conservative 57; Mismatches 159; Indels 74; Gaps 13;

QY 1 AGGGY-----W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYRSML 50  
 Db 27 ABGSYADLAEPDDDLVHVEGNIVDKYGKVIITCANWFGNCRERMLLDSYHSDIADI 86  
 QY 51 DOIKSLGYNTIRLPYSDIIL---KPGTMPNSINFYQMNQDQGLTSQVMDKIVAYAGQI 107  
 Db 87 ELVADKGINVVMPTATLLYAWSQGIYPPSTDTSYNPPALAGLSYELFNMLENFKRV 146  
 QY 108 GLRIILDRHDPDCSQS---ALWTSVSEATWSDLOALAQRYKNGNPTVVGFDLHNEPH 164  
 Db 147 GIKVILDVHSPETDNQGHNYPLVNTTITTEBFKAWVWVAERYKNDTTIIGFDLKHNEPH 206  
 QY 165 -----DPACWCGCDPSIDWRLAERAGNAVLSVNPILLIFVEGVQSY----- 206  
 Db 207 TNYGTWIKIAQSAIWDSDSHPNWKEVAETALALEVHPNVLIFVEGVEMYPKDGWDD 266  
 QY 207 -----NGDSY--WGNLQAGQYPPVL-NVFNRLVYSADHYATSVYPTQWFSDP 253  
 Db 267 EFTDTPMTNNDYXGNWGNLGRVGDYPINLGYQSOLVYSPHDYGPVIYEQDWFKGD 326  
 QY 254 TFPNN-----MPGIWNKXWGLYFNQNIAPVWMLGEGTTLQSTTDQTLKTLVQYLRP 305  
 Db 327 FTANDEQAKRILYQOCWRDNWAYIMERGISPLLLGEWGMTEG--GHPLDLNLKYLRC 384  
 QY 306 TAYQAGDS---FQWTFWSNWPSDGTGILKDD-----WQVDTVTKQGYLAPI 350  
 Db 385 MRDFILENKYKLHHTFWCINIDSADTGLFTRDEGTPPPGGRDLKWN--DNKYDNYLYPV 442

## RESULT 12

B82761  
 extracellular endoglucanase precursor XF0810 [imported] - Xylella fastidiosa (strain 9a5c  
 C:Species: Xylella fastidiosa  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: B82761  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc  
 Nature 406, 151-157, 2000  
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: B82515; MUID:20365717; PMID:10910347  
 A>Note: for a complete list of authors see reference number A59328 below

A:Accession: B82761  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-565 <SIM>  
 A:Cross-references: GB:AE003320; GB:AE003849; NID:g9105697; PIDN:AAE83620.1; GSPDB:GN0012  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al

F;20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 31.3%; Score 620.5; DB 2; Length 915;  
Best Local Similarity 35.5%; Pred. No. 1.9e-40;  
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRVSRMLDQIKSLGYNTIRLP 64  
DB WLVVSGNKIVDKGRPVWLTGYNWFGYNTGTVNFGVWSCKLDTLAEIANRGFNLLRVP 564  
QY 65 YSDIL---KPGTMRN-SINFYQNMQDLOGLTSLQWMDKIVAYAGQIGLRIILDRH--RP 118  
DB ISAEILNWSQGIYPKPNINY-VNPELECKNSLEVFDIVQTCKEVGLKIMLDIHSIKT 623  
QY 119 DCSGQ-SALWYTSVSEATWISDLQALQARYKGNPTVGVGFLHNEPHDP-----ACWG 170  
DB 624 DAMGHYPPWYDEKFTPEFYKACEMWITRYKNDTIIAFDLKNEVHGKFPWQDTTAKWD 683  
QY 171 CGDPSIDWRLAERAGNANVLNPNLLIFVEGVQSYNGD-----SYWGGNLL 217  
DB 684 NSTDINNWKYAAETCAKRIININPNLLIVIEGIEAYPKDDVTWTSKSSDYSTWGGNLL 743  
QY 218 OGAGQPPVVL-NVPNRLVYSAHDYATSVYPTQWFSPTFPNN--MPGIWNKNGYLFNQN 274  
DB 744 RGVKYPINLGKYNQKVVSPHYGSPVYQOPWFY-PGFTKESLLQDCWRPNWAYIMEEN 802  
QY 275 IAPVWLGEFTTILQSTTDTWLKTLVQYLRPTAQYGADSFQWTFWSNPDSGDTGGILKD 334  
DB 803 IAPLLIGEWGHLGDADNEKMKYLRDYII-----ENHIHTFWCFNANSBDTGLGVY 856  
QY 335 DWQTVDTVKDGYLAP 349  
DB 857 DFTTWDEKYSFLKP 871

## RESULT 8

S02711  
cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum  
N;Alternate names: endo-1,4-beta-glucanase  
C;Species: Caldocellum saccharolyticum  
C;Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 08-Oct-1999  
C;Accession: S02711  
R;Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.  
Nucleic Acids Res. 17, 439, 1989  
A;Title: Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo-  
A;Reference number: S02711; MUID:89098398; PMID:2789517  
A;Accession: S02711  
A;Molecule type: DNA  
A;Residues: 1-1039 <SAU>  
A;Cross-references: EMBL:X13602; NID:G40645; PIDN:CAA31936.1; PID:G40646  
C;Genetics:  
A;Gene: celB  
C;Function:  
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce-  
A;Pathway: cellulose degradation  
C;Superfamily: Streptomyces endo-1,4-beta-xylanase A homology  
C;Keywords: Glycosidase; hydrolase; polysaccharide degradation  
F;1-28/Domain: signal sequence #status predicted <SIG>  
F;25-1039/Product: cellulase #status predicted <MAR>  
F;72-373/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match 31.3%; Score 620.5; DB 2; Length 1039;  
Best Local Similarity 35.5%; Pred. No. 2.2e-40;  
Matches 133; Conservative 68; Mismatches 135; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRVSRMLDQIKSLGYNTIRLP 64  
DB 629 WLVVSGNKIVDKGRPVWLTGYNWFGYNTGTVNFGVWSCKLDTLAEIANRGFNLLRVP 688  
QY 65 YSDIL---KPGTMRN-SINFYQNMQDLOGLTSLQWMDKIVAYAGQIGLRIILDRH--RP 118  
DB 689 ISAEILNWSQGIYPKPNINY-VNPELECKNSLEVFDIVQTCKEVGLKIMLDIHSIKT 747

QY 119 DCSGQ-SALWYTSVSEATWISDLQALQARYKGNPTVGVGFLHNEPHDP-----ACWG 170  
DB 748 DAMGHYPPWYDEKFTPEFYKACEMWITRYKNDTIIAFDLKNEVHGKFPWQDTTAKWD 807  
QY 171 CGDPSIDWRLAERAGNANVLNPNLLIFVEGVQSYNGD-----SYWGGNLL 217  
DB 808 NSTDINNWKYAAETCAKRIININPNLLIVIEGIEAYPKDDVTWTSKSSDYSTWGGNLL 857  
QY 218 OGAGQPPVVL-NVPNRLVYSAHDYATSVYPTQWFSPTFPNN--MPGIWNKNGYLFNQN 274  
DB 868 RGVKYPINLGKYNQKVVSPHYGSPVYQOPWFY-PGFTKESLLQDCWRPNWAYIMEEN 926  
QY 275 IAPVWLGEFTTILQSTTDTWLKTLVQYLRPTAQYGADSFQWTFWSNPDSGDTGGILKD 334  
DB 927 IAPLLIGEWGHLGDADNEKMKYLRDYII-----ENHIHTFWCFNANSBDTGLGVY 980  
QY 335 DWQTVDTVKDGYLAP 349  
DB 981 DFTTWDEKYSFLKP 995

## RESULT 9

A40589  
cellulase (EC 3.2.1.4) - Clostridium thermocellum  
N;Alternate names: endo-1,4-beta-glucanase  
C;Species: Clostridium thermocellum  
C;Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 09-Jun-2000  
C;Accession: A40589; S31381  
J;Lenaire, M.; Beguin, P.  
J. Bacteriol. 175, 3353-3360, 1993  
A;Title: Nucleotide sequence of the celG gene of Clostridium thermocellum and characteri-  
A;Reference number: A40589; MUID:93273701; PMID:8501039  
A;Accession: A40589  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-566 <LSW>  
A;Cross-references: GB:X69390; NID:G40677; PIDN:CAA49187.1; PID:G40678  
C;Genetics:  
A;Gene: celG  
C;Function:  
A;Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce-  
A;Pathway: cellulose degradation  
C;Superfamily: cellulase B; Clostridium cellulase repeat homology  
C;Keywords: Glycosidase; hydrolase; polysaccharide degradation  
F;503-526/Domain: Clostridium cellulase repeat homology <CCR>  
F;536-559/Domain: Clostridium cellulase repeat homology <CCR>

Query Match 30.1%; Score 596.5; DB 2; Length 566;  
Best Local Similarity 34.2%; Pred. No. 7.4e-39;  
Matches 139; Conservative 55; Mismatches 143; Indels 69; Gaps 15;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRVSRMLDQIKSLGYNTIRLP 64  
DB 44 WLHCKGNKIYDMYGNVWLTGANWFGNFCSECFHG-AWYDVKTILTSLADRGINLLRIP 102  
QY 65 YSDIL---KPG-----TMPNSINFYQNMQDLOG-----LTSQVMDKIVAYAGQI 107  
DB 103 ISTEILSYMWLGKPNPSSVTASNNPYPHVNPDPYDPTDDVKNSEIFDINGYCKEL 162  
QY 108 GLRIILDRHDPDC-SGQS-ALWY---TSS---VSEATWISDLQALQARYKGNPTVGVGFD 158  
DB 163 GIKVMDIHSPPDANSGHNYELWYKETSCTGVVTTKMWIDTLVWLAKYKNDDTIAFD 222  
QY 159 LHNEPHDP-----ACWGGDPSIDWRLAERAGNANVLNPNLLIFVEGVQSY- 206  
DB 223 LKNEPHGKRGYAEVPKLLAKWDSNDENNWKYAAETCAKRIILEVNPVKLIVIEGVQYP 282  
QY 207 -----NGD-----SYWGGNLSQAGQYPPVVLNPN-RLVTSAGHDYATSV 244  
DB 283 KTEGYTYDTPDIWCATGADSPWYSAWWGNLGRGVKDYPIDLGPLNSQIVYSPHDYGPSV 342  
QY 245 YPQTFW-SDTFPPNMPGIIWNKNGYLFNQNIAFWLGEFTTILQSTTDTWLKTLVQYL 303

QY 228 NVNRLVYSAHDYATSYVQTPWSDPTFFNNMPCGNKXWGLFNQNIAPVWLGEGTTL 287  
Db 268 --PRGKVVYSHVGVSPYMMDYFKSPDPFNNMPLIWEHFGYLTDLNLYTLV-IGWGGNY 325  
QY 288 QSTTDTWLKTLVQLRPTAQYAGDSFQWTFWSNPDGDTGGILKDDWQTVDTVK 343  
Db 326 EG-LDKVQDAFVKWLKKIY-----NFFWCLNPESGDTGGILKDDWQTVNWEK 375  
  
RESULT 5  
E82759  
endo-1,4-beta-glucanase XF0818 [imported] - Xylella fastidiosa (strain 9a5c)  
C:Species: Xylella fastidiosa  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: E82759  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
Nature 406, 151-157, 2000  
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A:Reference number: A82515; MURD:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82759  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-592 <SIM>  
A:Cross-references: GB:A8003921; GB:A8003849; NID:g9105710; PIDN:AAF83628.1; GSPDB:GN001  
A:Experimental source: strain 9a5c  
B:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laigh  
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
Rodrigues, V.; Rosa, A.J.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaak  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF0818  
  
Query Match 34.5%; Score 683.5; DB 2; Length 592;  
Best Local Similarity 42.6%; Pred. No. 1.2e-45;  
Matches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;  
  
QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSMLDOIKSLGYNITRLP 64  
Db 26 YSHGKVIDDKGN-QIOLKGGWFGFETTNVHVHGLWLNKWEFTIQIOSMGLNVLRLP 84  
QY 65 YSDILKPGTWPNSINFYOMNQDQLGHTSLQVMDKIVAYAGQIGLRIILDRHPDCSGS 124  
Db 85 FCPASLNSNTSPSSID-YNRNPDQLGLSLQIMDKVVKELSDRGIVLVMDHHPDCAAIS 143  
QY 125 ALWTSSSEATWISDLOALQAKYKGNPTVVGDLNHPDHPACGGDSIDWRLAAER 184  
Db 144 ELWTDYSSEKQWIDDLRFVAHRYANYPGVIGLSDVRNPHGRATWGTGDPKTDWNTAVEH 203  
QY 185 AGNVLNVNPLLIIFVGVQSYNGDS----YWGGNQLQAGQYPPVNLNVP-NRLVYSAHD 239  
Db 204 AAAAILEAPKWLIGVEGIGENPSCSTIGHFAGENLEPMDCTP-LKVPADHLLIIPHV 261  
QY 240 YATSVYQTFWSDPTFFNNMPCGNKXWGLFNQNIAPVWLGEGTTL--QSTTDTWLK 297  
Db 262 YGPDVYVQYFNSPDPFNNMAAIWDKHFHFAKAGYA-NAIGFPGKYGKGGPRDVAQN 320  
QY 298 TLVQLRPTAQYAGDSFQWTFWSNPDGDTGGILKDDWQTV 339  
Db 321 ALVDYL---ISIGVTDIF---YMNANSSDTGGLFEDDNNHV 357  
  
RESULT 6

E97012  
probable non-processive endoglucanase family 5, secreted, Cella homolog secreted, dockerrir  
C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: E97012  
R:Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A:Reference number: A56900; MUID:21359325; PMID:21359325  
A:Accession: E97012  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-482 <KUR>  
A:Cross-references: GB:A3001437; PIDN:AAK78888.1; PID:g15023812; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC0912  
  
Query Match 34.0%; Score 674; DB 2; Length 482;  
Best Local Similarity 37.6%; Pred. No. 5.1e-45;  
Matches 140; Conservative 71; Mismatches 127; Indels 34; Gaps 9;  
  
QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSMLDOIKSLGYNITRLP 64  
Db 24 YLHSDGSKLLDDYGNQVEMTGIAWFGLETPNYCFHGLWANRLDNLINVDNGNTLRVP 83  
QY 65 YSDIL---KPGT--MPNSINFYOMNQDQLGHTSLQVMDKIVAYAGQIGLRIILDRHPD 119  
Db 84 LSELVNWQRQGVYTPPSINDY-ISPCLKGQNSLIQLDDVIAYSKKVGVKVMWDMHRIE 142  
QY 120 CSGSALWYTSVSSEATWISDLOALQAKYKGNPTVVGDLNHPDHPAC-----NGCG 172  
Db 143 SGGQTATWTSKYTTDDYEKCQWVLADRYKNDTVAADIFNEPHGKAYRAETSAKNDT 202  
QY 173 DSIIDWRLAABRAGNVLNVNPLLIIFVGVQSY-----NGDSY---WGGNQLQ 219  
Db 203 TDENWRVEAEKVGKKLIDINPKMLIVVEGVETPKGTAGAGSTNPDYGGWGGNLRG 262  
QY 220 AGQYPPVL-NVFNRLVYSAHDYATSYVQTFW-SDPTFFNNMPCGNKXWGLFNQNIAP 277  
Db 263 VKDYPDLAPYKQVYSPHDPGVGSDQTFWFGDFTQSLINDIWRSPWFYQEKNIAP 322  
QY 278 VWLGFETLQSTTDTQTLKTLVQLRPTAQYAGDSFQWTFWSNPDGDTGGILKDDWQ 337  
Db 323 LLIGEWGNNMDGGKNEQWMTDVKLI-----SDNKNMHTFWCLNANSAGDTGGILEYDFK 376  
  
QY 338 TVDVTVKDGYLAP 349  
Db 377 FIDTKLALVQP 388  
  
RESULT 7  
A43802  
cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum se  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Caldocellum saccharolyticum  
C:Date: 30-Jan-1993 #sequence\_revision 30-Sep-1993 #text\_change 10-Jul-1998  
C:Accession: A43802  
R:Saul, D.J.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.  
Appl. Environ. Microbiol. 56, 3117-3124, 1990  
A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile "C  
A:Reference number: A43802; MUID:91136262; PMID:2126700  
A:Accession: A43802  
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
A:Molecule type: DNA  
A:Residues: 1-915 <SAU>  
A:Cross-references: EMBL:X13602  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
A:Pathway: cellulose degradation  
C:Superfamily: Streptomyses endo-1,4-beta-xylosanase A homology  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation



JH0158  
cellulase (EC 3.2.1.4) precursor - Xanthomonas campestris pv. campestris  
N/Alternate names: endo-1,4-beta-glucanase; extracellular endoglucanase  
C/Species: Xanthomonas campestris pv. campestris  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 24-Sep-1999  
C/Accession: JH0158  
R/Gough, C.L.; Dow, J.M.; Keen, J.; Henrissat, B.; Daniels, M.J.  
Gene 89, 53-59, 1990  
A/Title: Nucleotide sequence of the engXCA gene encoding the major endoglucanase of Xanthomonas campestris pv. campestris  
A/Reference number: JH0158; MUID:90323605; PMID:2373365  
A/Accession: JH0158  
A/Molecule type: DNA  
A/Residues: 1-493 <GOU>  
A/Cross-references: GB:M32700; NID:g153397; PIDN:AA227612.1; PID:g153398  
C/Comment: Endoglucanase plays a minor role in the early stages of pathogenicity of Xanthomonas campestris pv. campestris  
C/Genetics:  
A/Gene: engXCA  
C/Function:  
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellobiose  
A/Pathway: cellulose degradation  
C/Superfamily: Xanthomonas campestris cellulase  
C/Keywords: glycosidase; hydrolase; polysaccharide degradation  
F/1-25/Domain: signal sequence #status predicted <SIG>  
F/26-474/Product: cellulase #status predicted <MAT>  
F/375-399/Region: proline/threonine-rich

Query Match 40.0%; Score 792.5; DB 1; Length 493;  
Best Local Similarity 46.9%; Pred. No. 2.9e-54;  
Matches 161; Conservative 59; Mismatches 104; Indels 19; Gaps 9;

QY 5 YNHTSGREILDANNVPVRIAGINWFGFCNYYVHGLSRDYSRMLDQIKSLGYNTIRLP 64  
DB 26 YSINNSRQIVDDSGKVVQKLVGNVFGFETGNHVHGLWARKMDLVQMGCGFNAVRLP 85

QY 65 YSDILKPGTMPNSINFYQMNQDQLGTSIQVMDKIVAYAGQIGRIILDRHRPCSGS 124  
DB 86 FCPATLRSDTPASID-YSRNADLGTSILQIDKVIABFNARGMYVLDDHTPTDCAGIS 144

QY 125 ALMYTSSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184  
DB 145 ELMYTSYTEAQLADLRVFNKVPVYGLDLKXNEPHGATWGTGNAATDNKAER 204

QY 185 AGNAVLNVNPLLIIFVEGVQ-----SYNGDSYWMGNLQAGQYPVVLPVNP-NRLVYSAH 238  
DB 205 GSAAVLAVAPKWLIAVEGITDNPVCGTNG-GIFWGNLQPLACTP-LINIPANRULLAPH 261

QY 239 DYATSVYPTQWFSDDPTFPNNPGIWNKNGWYLFNQNIAPVWLGEFTTL--QSTTDQTL 296  
DB 262 VYGPVDFVQSYFNDSPNPNMPLAERHFQCFAGTH--ALLLGEFGKYGECDARDKTWQ 319

QY 297 KTLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQTV 339  
DB 320 DALVKYLR---SKGIN--QGFYWSNPNDSGDTGGILRDDWTSV 357

RESULT 3  
E71059  
probable endo-1,4-beta-glucanase - Pyrococcus horikoshii  
C/Species: Pyrococcus horikoshii  
C/Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
C/Accession: E71059  
R/Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Hailkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A/Reference number: A71000; MUID:98344137; PMID:9679194  
A/Accession: E71059  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-458 <KAW>  
A/Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30271.1; PID:g3257588  
A/Experimental source: strain OT3  
A/Note: this accession replaces an interim accession for a sequence replaced by GenBank

C/Genetics:  
A/Gene: PH171  
C/Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 37.9%; Score 751; DB 2; Length 458;  
Best Local Similarity 45.0%; Pred. No. 4.6e-51;  
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWFGFCNYYVHGLSRDYSRMLDQIKSLGYNTIRLPYS 67  
DB 54 TSGEE-----TPHLFGVNVWFGFETPNVHGLWARKMDLVQIKSLGFNAIRLPFCT 107

QY 68 DILKPGTMPNSINFYQMNQDQLGTSIQVMDKIVAYAGQIGRIILDRHRPCSGQSALW 127  
DB 108 ESKAPGQPTGID-YSKNPLRGLDLSQIKWEKLIKAGDLGIFVLDDYHGRIGTHIEPLW 166

QY 128 YTSSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175  
DB 167 YTEDFSEEDFINTWIEVAKRFKYNVIGADLKNPESHVTSPPAAVYTDGTGATWGMGNA 226

QY 176 IDWRLAERAGNAVLSVNPPLLIIFVEGVQSYN-----GDSYWMGNLQAGQYPVV 226  
DB 227 TDNLAARERIGKAILKVAHPHLIFVEGTQTNPKTSSYKWGNWMMWGMMAVKYIPV- 285

QY 227 LNVP-NRLVYSAHDYATSVYPTQWFSDDPT-FPNNMPGIMNKNWGYLFNQNIAPVWLGEFG 284  
DB 286 -NLPRNKLVSYPHYGPDVYNQPVFGPAKGFDPNLPDIWYHFGYVVKLELGSYVIGFEG 344

QY 285 TTL---QSTTDQTLVQLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDQTV 339  
DB 345 GYGHGSDPRDVIWQNKLVDMW--IENKFCDF--YWSNPNDSGDTGGILQDDWTII 397

RESULT 4  
E75142  
endoglucanase PAB0632 - Pyrococcus abyssi (strain Orsay)  
C/Species: Pyrococcus abyssi  
C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
C/Accession: E75142  
R/anonymus, Genoscope  
A/Description: submitted to the EMBL Data Library, July 1999  
A/Reference number: A75001  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-514 <KAW>  
A/Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49854.1; PID:g545836;  
A/Experimental source: strain Orsay  
C/Genetics:  
A/Gene: celB-like; PAB0632  
C/Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 35.0%; Score 694.5; DB 2; Length 514;  
Best Local Similarity 41.9%; Pred. No. 1.4e-46;  
Matches 149; Conservative 57; Mismatches 121; Indels 29; Gaps 10;

QY 7 HTSGREILDANNV-----PVRIAGINWFGFETPNVHGLSRDYSRMLDQIKSLGYNT 60  
DB 30 YTAENGIIIFVQVNTTGEKKPLYLHGVSFGLXDVVYGLDKENWKDILKDKVRELGFNA 89

QY 61 IRLPYSDILKPGTMPN--SINFYQMNQDQLGTSIQVMDKIVAYAGQIGRIILDRHRP 118  
DB 90 IRLPCEISIRPDRTPSPERIN-YELNPD LKNLISLEIMEXIIIEYANSIGLYILLDYHRI 148

QY 119 DCSGOSALWTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDW 178  
DB 149 GCEETEPLWYNTENYSEQYIKDWIFLAKRFKYPNVIGADIKNEPHGAGTGDGR--DF 207

QY 179 RLAAERAGNAVLSVNPPLLIIFVEGVQ-----SYNGDSYW--WGNLQAGQYPVV 227  
DB 208 RLFAEKVGREILKVAHPHLIFVEGTQYTHVNPIDIEIKKGNWTFWGNLWGVKDYVRL 267

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:59 ; Search time 12.3333 Seconds

(without alignments)

2792.154 Million cell updates/sec

Title: US-09-997-504A-12

Perfect score: 1982

Sequence: 1 AGGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1038	52.4	397	2 A35136	cellulase (EC 3.2.
2	792.5	40.0	493	1 JH0158	cellulase (EC 3.2.
3	751	37.9	458	2 E71059	probable endo-1,4-
4	694.5	35.0	514	2 E75142	endoglucanase PAB0
5	683.5	34.5	592	2 E82759	endo-1,4-beta-gluc
6	674	34.0	482	2 E97012	probable non-proce
7	620.5	31.3	915	2 A43802	cellulase (EC 3.2.
8	620.5	31.3	1039	2 S02711	cellulase (EC 3.2.
9	596.5	30.1	566	2 A40589	cellulase (EC 3.2.
10	541.5	27.3	747	2 B47093	cellulase (EC 3.2.
11	528	26.6	563	1 CZCLEM	cellulase (EC 3.2.
12	328	16.5	565	2 E82761	extracellular endo
13	290.5	14.7	748	2 S19652	cellodextrinase C
14	252.5	12.7	522	2 E86255	protein F319.15 [
15	214	10.8	526	2 T51476	hypothetical prote
16	204.5	10.3	488	2 T51502	hypothetical prote
17	167	8.4	426	2 A42649	cellulase (EC 3.2.
18	154.5	7.8	722	2 H96986	endo-1,4-beta gluc
19	151.5	7.6	356	2 G82523	endo-1,4-beta-gluc
20	149.5	7.5	329	2 J72216	endoglucanase - Th
21	149.5	7.5	516	2 E01334	cellulase (EC 3.2.
22	147.5	7.4	814	1 CZCLEM	cellulase (EC 3.2.
23	147	7.4	1012	2 E97326	endoglucanase fami
24	145.5	7.3	406	2 A43722	cellulase (EC 3.2.
25	140.5	7.1	669	2 D72278	endo-1,4-beta-mann
26	138.5	7.0	500	2 S22458	cellulase (EC 3.2.
27	137	6.9	370	2 G97001	endoglucanase fami
28	134	6.8	409	2 S12018	endoglucanase B -
29	134	6.8	584	2 JQ1229	cellulase (EC 3.2.

#### ALIGNMENTS

##### RESULT 1

A35136  
cellulase (EC 3.2.1.4) - Bacillus polymyxa  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Bacillus polymyxa  
C>Date: 03-Aug-1990 #sequence\_revision 03-Aug-1990 #text\_change 24-Sep-1998  
C/Accession: A35136  
R:Baier, S.D.; Johnson, D.A.; Seligy, V.L.  
J. Bacteriol. 172, 1576-1586, 1990  
A:Title: Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase S  
A:Reference number: A35136; MUID:90170877; PMID:2307659  
A:Accession: A35136  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-397 <BAI>  
A:Cross-references: GB:M33791; GB:M33840  
C:Function:  
A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
A:Pathway: cellulose degradation  
C:Superfamily: Xanthomonas campestris cellulase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 52.4%; Score 1038; DB 2; Length 397;  
Best Local Similarity 54.2%; Pred No. 1.3e-73;  
Matches 193; Conservative 44; Mismatches 107; Indels 12; Gaps 4;

QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMLEDOIKSLGYNTRL 63  
DB 37 GYHTQGNKIVDESGKEAFAFNGLNWFGLETPNYTLHGLWSRMDMLDQVKKEGYNLRL 96  
QY 64 PYSDILKFGMPINSEYQMNQDLGLTSLQVMDKIVAYAGIGLRIILDRHPCDSGQ 123  
DB 97 PYSNOLFQSSSRPDSID-YHKNPDLVGLNPQIQIMDKLLEKAGQRIQIILDRHPCSGGQ 155  
QY 124 SALWTSSVSRAETWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWGCGPDSIDWRLAAE 183  
DB 156 SELWYTSQYPSRWISDWKMLADRYKNNFTVIGADLHNEPHGQASWGSGTGNASTDWRLLAAQ 215  
QY 184 RAGNAVLSVNPMLLIFFVEGVQ---SYNGDSYWGNGNLOGAGQYPVVLNPNRLVYSAHDY 240  
DB 216 RAGNAVLSVNPMLLIFFVEGVQ---SYNGDSYWGNGNLOGAGQYPVVLNPNRVVYSPHDY 275  
QY 241 ATSVYPTQWFSFDPPTPPNNMPPGWNKNGYLEFNQNTAPVWLGEFG--TFLQSTTDTQTLTKT 298  
DB 276 GFGVSSQWQFNDPAFPSNLPALWDQWGYISKQNLAPLVGFEFGGRNVLDSSPEGKQNA 335  
QY 299 LVQYLRPTAQVGADESFQWTFWSWNPDSQDTGGILKDDWQTVDTVDKGVLYAPIKSSI 354  
DB 336 LVHYI-----GANNLYFTYSLNPNSGDTGGLLDWTTWNRPKQDMLGRMKPV 385

##### RESULT 2

This Page Blank (uspto)

```
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9262
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9262

Query Match          32.9%; Score 651.5; DB 15; Length 421;
Best Local Similarity 39.9%; Pred. No. 3.4e-55;
Matches 141; Conservative 60; Mismatches 123; Indels 29; Gaps 12;

Qy 5 YWHTSGREILDANNVPVRIAGINWFGPETCNVYVHGLRSDYRSMLDQIKSLGYNTIRLP 64
Db 26 YSISHGKVVDKGN-QIQLRGVNWFGPETGHDVYVNGLVARNWKEFITQLQGMGFNAIRLP 84
Qy 65 YSDDILKPGTWPNISINFYQMNQDLQGLTSLQVMDKIVAVAGQIGLRILIDRHRPDCSGQS 124
Db 85 FCPANLNSNTPSSID-YSRNPDQLGLSSQILDKVVKELSDRRMYVLLDHRPDCSAIS 143
Qy 125 ALWYTSSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAER 184
Db 144 ELWHTDSVSEKQWIDDLRFVAHRVANVPVGVLGLDVKNPHGRATWGTGDFQTDWNTAVEH 203
Qy 185 AGNAVLSVNFNLLIFVSGVQSYNGDS-----YWGGNLQAGAGQYPVVLNYP-NELVY 235
Db 204 AAAAILLEAAPKWKVIGIGI---GENPICSTIGHFWGENLEPMDCPT--LKYPADHLLL 257
Qy 236 SAHDYATSVYPTWFSPTFPNNMGIWNKNWGYLFNQNIAPVWLGEFGTTL--QSTTDQ 293
Db 258 MPHVGPDVYVQYFNSGDFPNNMAAIWDKHFHFAKAGYA-MAIGFEGKYGEGDPRDI 316
Qy 294 TWLKTLYQYLRPTAQYG-ADSFQWTFWSWNPDSGDTGILKDDWQTV--DTVK 343
Db 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSGDTGGMVGNWNTTPRDDKVK 363
```

Search completed: August 2, 2004, 16:40:49  
Job time : 42.3333 secs

[illegible]

RESULT 13  
 US-10-369-493-17534  
 ; Sequence 17534, Application US/10369493  
 ; Publication No. US2003023675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
 ; FILE REFERENCE: 38-10 (52052) B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 17534  
 ; LENGTH: 592  
 ; TYPE: PRT  
 ; ORGANISM: *Xylella fastidiosa*  
 US-10-369-493-17534

Query Match	34.5%	Score	683.5	DB	15	Length	S92
Best Local Similarity	42.6%	Pred.	No. 3.9e-58				
Matches	146	Conservative	50	Mismatches	128	Indels	19
Gaps	10						
QY	5	YWHTSGREILDANNVPRKIAGINWFGFEFTCNVYVHGLRSRDYRSMGLQIKSLGYNTIRLP	64				
DB	26	YSISHGKVIDDKGN-QTCLKGGSWFGFETTNVHVLWTRNWKKEFIQTIQSMGLNAVRLP	84				
QY	65	YSDIILKEPTWPNISINFQVMQDLOGLTSLQWMDKIYAYACQIGRIILDBRHPPDCSGQS	124				
DB	85	FCPASLNSNTSPSSID-YNRNPDIOGLSSLOIMDKWKELSDRGIVYLMDDHTDPDCAAIS	143				
QY	125	ALWYTSYSEATWISDLQALAQRYKGNPTVYVGFDLHNEPHDPACWCGDPSIDWELAAER	184				
DB	144	ELWYTSYSEKQWIDDLRFVHRYANVPVGIVGLDVKNEPHGRATWGTGDPKTDWNTAVEH	203				
QY	185	AGNAVLSVNPMLLTVEGVQSYNGDS-----TWGGNLOGAQCYQVYLVNVP-NRLVYSAHID	239				
DB	204	AAAAIILEAAPKWLIVGEVIGENPSCSSTIGHFWGENTLEPMDCTP-LKVPADHLLILPHV	261				
QY	240	YATSYVPOTWTSDDPTFPNNMFGIMNKNWGYILFNQNIAPWILGERGTTIL--QSTTDQTWLK	297				

```

Db      262 YGPDVYVQVPFNSPDFPNMAATWDFHGFHAKAGYA-MAIGFGKGKYGEGDPRDVAMQN 320
          :|: |:: ||||| ::||:: :|: |:|:
Qy      298 TLUVXLRTAQYG-ADSFQWTFFSWNPDSDGTGGILKKDDWQT V 339
          :|: |:: ||||| ::||:: :|: |:|:
Db      321 ALVDYL---ISGVTDTF---YWSINANSSDTGGLFRDDWNHV 357
          :|: |:: ||||| ::||:: :|: |:|:

RESULT 14
US-10-369--493-9473
; Sequence 9473, Application US/10369493
; Publication NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
; OF ENRICHED PRODUCTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIORITY CLAIMING DATE: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9473
; LENGTH: 369
; TYPE: PRP
; ORGANISM: Xylella fastidiosa
US-10-369-493-9473

```

	Query Match	33.1%	Score 656.5	DB 15	Length 369
	Best Local Similarity	40.5%	Pred. No. 9.1e-56		
	Matches	143	Conservative 56	Mismatches 125	Indels 29
					Gaps 12
Qy	5	YWHTSGREILDANNVPYRIAGINWFGPETHCNVYVHGLRSRDYRSMLDQIKSIGVNTIRLP	64		
Db	26	YSISHGKVVDDKGN-QTQLRGVNVFGPETHVHVGWGLWARKWKEFITQCGWGFNAIRLP	84		
Qy	65	YSDDILKPGTWPNSINFPQMNDIQGLTSLQWMDKIYAYAGIQIGLRIILDRHRPDCSQS	124		
Db	85	FCPANLNSNTSPSSID-YSRNPDQLGLSSQLDKVVKELSDRRMYVLLDHRSDCSAIS	143		
Qy	125	ALWYTSSVSEATWISDLOALAQRYKGNPTVVGFIDHNEPHDPACWCGDPSIDRWLAER	184		
Db	144	ELWYTDYSGEKQWIDDLRFVAHRYTNVPGVIGLVQKNEPFGHRTWGTGDPKTDMTVAEH	203		
Qy	185	AGNAVLSVNPMLLIFFVSGVQSVNGDS-----YVMGGNIQAGQYVPWLNVV-NRLVY	235		
Db	204	AAAAILEAPKWLIVGSEI-----GENPICSTIGHFGENLEPMDCTP--LKVPANHLLL	257		
Qy	236	SAHDYATSVYPQTWFSPTFPNNMPGWNKMWGLFNQNIAPFVIMLGEFTLL--QSTTDQ	293		
Db	258	MPHYGPDVYVQYPNSPDPFNNNYAAIWDQHFGRFAKAGYA-MAIGEPGGKYGEGDPRDI	316		
Qy	294	TWLKTLVQYLRPTTAQYG-ADSFQWTFWSNPDSDGTGGILKDDKQTV--DVTK	343		
Db	317	AWONAFVDYL---ISIGVTDFAF---YVAAONQSGDTGMVGNDMTTPRDDKVK	363		

RESULT 15  
US-10-369-493-9262  
; Sequence 9262, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Stater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

```
Db 167 YTEDFSEEDFINTWIEVAKRFGKYNNVIGADLKNEPHSVTSPPAAVYTDGTGATWGMGNDP 226
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYMWGNLQAGQYPPV 226
Db 227 TDMNLAERIGKAILKVAPHWLIIFVEGTQNTPKTSSYKWNWGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFS DPT-PPNNMGINNKNWGYLFNQNIAPVWLGEGF 284
Db 286 -NLPRNKLVSHPYVGPVYNQPYFGPAKGFPPDNLPIWYHFGYVLELGYSVWIGEGF 344
Qy 285 TTL---QSTTDQWLKTLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 345 GKYHGSDPRDVIWONKLVDMW--IENKFCDF--YWSNPNDSGDTGGILQDDWTTI 397

RESULT 10
US-10-369-493-1281
; Sequence 1281, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1281

Query Match 37.9%; Score 751; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.1e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPVRIAGINWFGPETHCNVYVHGLRSRDRYSMLDQIKSLGYNTIRLPYSD 67
Db 54 TSSEE-----TPHILFGVNMFGFETPNHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQSALW 127
Db 108 ESKVPGTQPTGID-YSKNPDLRLGLDSLQIMEKLIKAGDGI FVLDDYHRI GCTHIEPLW 166
Qy 128 YTSVSEATWISDLOALQAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEDFSEEDFINTWIEVAKRFGKYNNVIGADLKNEPHSVTSPPAAVYTDGTGATWGMGNDP 226
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYMWGNLQAGQYPPV 226
Db 227 TDMNLAERIGKAILKVAPHWLIIFVEGTQNTPKTSSYKWNWGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFS DPT-PPNNMGINNKNWGYLFNQNIAPVWLGEGF 284
Db 286 -NLPRNKLVSHPYVGPVYNQPYFGPAKGFPPDNLPIWYHFGYVLELGYSVWIGEGF 344
Qy 285 TTL---QSTTDQWLKTLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 345 GKYHGSDPRDVIWONKLVDMW--IENKFCDF--YWSNPNDSGDTGGILQDDWTTI 397

RESULT 11
US-10-369-493-20347
; Sequence 20347, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1281

Query Match 37.9%; Score 751; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.1e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPVRIAGINWFGPETHCNVYVHGLRSRDRYSMLDQIKSLGYNTIRLPYSD 67
Db 54 TSSEE-----TPHILFGVNMFGFETPNHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQSALW 127
Db 108 ESKVPGTQPTGID-YSKNPDLRLGLDSLQIMEKLIKAGDGI FVLDDYHRI GCTHIEPLW 166
Qy 128 YTSVSEATWISDLOALQAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175
Db 167 YTEDFSEEDFINTWIEVAKRFGKYNNVIGADLKNEPHSVTSPPAAVYTDGTGATWGMGNDP 226
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYMWGNLQAGQYPPV 226
Db 227 TDMNLAERIGKAILKVAPHWLIIFVEGTQNTPKTSSYKWNWGNLMAVKDYPV- 285
Qy 227 LNVP-NRLVYSAHDYATSVYPTWFS DPT-PPNNMGINNKNWGYLFNQNIAPVWLGEGF 284
Db 286 -NLPRNKLVSHPYVGPVYNQPYFGPAKGFPPDNLPIWYHFGYVLELGYSVWIGEGF 344
Qy 285 TTL---QSTTDQWLKTLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGILKDDWQTV 339
Db 345 GKYHGSDPRDVIWONKLVDMW--IENKFCDF--YWSNPNDSGDTGGILQDDWTTI 397
```

; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: US 60/279,493  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 69  
; SEQ ID NO 50  
; LENGTH: 535  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-09-927-827-10

Query Match 39.6%; Score 785.5; DB 10; Length 535;  
Best Local Similarity 46.6%; Pred. No. 36-68;  
Matches 160; Conservative 59; Mismatches 105; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVVRVIRAGINWFGFETCNVYVHGLRSRDRYRMLDQIKSLGYNTIRLP 64  
DB 77 YSINRSQVDSQKVVQKGVNFGFETCNVYVHGLRWKMDVQVQGLGFAVRLP 136  
QY 65 YSDILKPGTMPSINFYQNMQLQGLTSLOVMDKIVAYAGQIGRIILDRHRPDCSQS 124  
DB 137 FCPATLRSDTPEASID-YERNADLQGLTSLOVMDKIVAYAGQIGRIILDRHRPDCAGIS 195  
QY 125 ALWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPAACHGCGDPSIDWELAAR 184  
DB 196 ELWYTSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHDPAACHGCGDPSIDWELAAR 255  
QY 185 AGNAVLSVNPVNLIFVEGVQ-----SYNGDSYVWGNLQAGQYFVVLNVP-NRLVYSAH 238  
DB 256 GSAAVLAPKWLIAVEGIDTNPVCSGTNG-GIFWGNLQPLACTP--LNIPANRLLLAPH 312  
QY 239 DVATSVYPTWSDPTFPNMGIWKNWGYLFNQNIAPVWLGFGFTTL--OSTTDTWL 296  
DB 313 VYGPVFGVSYFNSFNPNMFAIWERHFGQFAGTH--ALLIGBFDGKYGEADSKTWQ 370  
QY 297 KTLVOYLRTAQYAGDSFQWTFWSPNPDSDGTGGILKDDWQTV 339  
DB 371 DALVKYLR---SKGIN--QGFYKSNPNPSGDTGGILRDDWTSV 408

RESULT 8  
US-09-888-224-2  
; Sequence 2, Application US/09888224  
; Patent No. US20020120118A1  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CECCHIA, BYRNE, BAIN, GILFILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/888,224  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/430,669  
; FILING DATE: 28-Oct-1999  
; APPLICATION NUMBER: US/08/651,572  
; FILING DATE: 22-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Herron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS: Unknown  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-09-888-224-2

Query Match 38.2%; Score 758; DB 9; Length 553;  
Best Local Similarity 45.2%; Pred. No. 1.6e-65;  
Matches 160; Conservative 50; Mismatches 108; Indels 36; Gaps 11;

QY 13 ILDANNVVRVIRAGINWFGFETCNVYVHGLRSRDRYRMLDQIKSLGYNTIRLPYSDDLKP 72  
DB 4 VATGEETPIHLFGVNWFGFETPNVYVHGLMSRWEDMLQIKSLGFNAIRLPCTQSVPK 63  
QY 73 GTVPNSINFYQNMQLQGLTSLOVMDKIVAYAGQIGRIILDRHRPDCSQSALWTSSV 132  
DB 64 GTMPTAID-YAKNPDLQGLDSVQIMEKIIKAGDLGIFVLVDYHRIGCNFIEPLWYDTSF 122  
QY 133 SEATWISDQALAQRYKGNPTVVGFDLHNEPHD-----ACWCGDPSIDWRL 180  
DB 123 SEQDYINTWYEAQRFQKYNWVIGADLKNPHSSSPAPAAAYTDGSGATWGMGNATDNL 182  
QY 181 AAERAGNAVLSVNPVNLIFVEGVQ-----SYN-GDSYVWGNLQAGQYFVVLNVP- 230  
DB 183 AAERIGRAILEVAPQWVIFVEGTQFTTFEIDGRYKWHNAWGNLGMVYKYPV--NLPR 240  
QY 231 NRLVYSADHATSVYPTWFSDB--TPNMPGIWKNWGYLFNQNIAPVWLGFGFTTL- 287  
DB 241 DKVYSPQVYGVSEYDQYF-DRGEGFPDNLPEIWHYHFGYVVKLDLGYPVVIGFSGKYG 299  
QY 288 --QSTTDTWLKTLVOYLRTAQYAGDSFQWTFWSPNPDSDGTGGILKDDWQTV 339  
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDFP---YVSNPNPSGDTGGILKDDWTII 348

RESULT 9  
US-10-293-344A-2  
; Sequence 2, Application US/10293344A  
; Publication No. US20030129723A1  
; GENERAL INFORMATION:  
; APPLICANT: RAKUTO KASEI INDUSTRIAL CO., LTD. and  
; APPLICANT: National Institute of Advanced Industrial Science and Technology  
; TITLE OF INVENTION: Thermophilic endoglucanase  
; FILE REFERENCE: P017123A  
; CURRENT APPLICATION NUMBER: US/10/293,344A  
; CURRENT FILING DATE: 2003-03-17  
; NUMBER OF SEQ ID NOS: 4  
; SEQ ID NO 2  
; LENGTH: 458  
; TYPE: PRT  
; ORGANISM: Pyrococcus horikoshii  
; FEATURE:  
US-10-293-344A-2

Query Match 37.9%; Score 751; DB 14; Length 458;  
Best Local Similarity 45.0%; Pred. No. 6.1e-65;  
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVVRVIRAGINWFGFETCNVYVHGLRSRDRYRMLDQIKSLGYNTIRLPYS 67  
DB 54 TSSEE-----TPHLEGVNWFGFETPNVYVHGLWKNWEDMLQIKSLGFNAIRLPCT 107  
QY 68 DILKPGTMPSINFYQNMQLQGLTSLOVMDKIVAYAGQIGRIILDRHRPDCSQSALW 127  
DB 108 ESKVPGTQPIGID-YSKNPDLRGILDSLQIMEKIIKAGDLGIFVLVDYHRIGTHIEPLW 166  
QY 128 YTSSVSEATWISDQALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175

QY	5	YWHTSGREITLDANNVPVRIAGINNFGFETCNVYVHGLRSDYRSMDDQIKSLGYNTIRLP	64
Db	26	YGINNSRQIVDDSGKVQLKGVNVFGFETGNVYVHGLWARNWKDMIVQMGLGFNAVRLP	85
QY	65	YSDDILKCGTWPNISNFMQNDIQLGLTSLQWDXKIVAYAGICLRIILDRHRPDCSQS	124
Db	86	FCPATLRSDTWPASID-YSRNADLQGLTSLQILDKVIAEFNARGMTVLLDHHHTPPDCAGLS	144
QY	125	ALWYTSYSSSEATWISDLQALAQRVKGNPVTVVGFDLHNEPHDPACWCGDPSIDWRLAAER	184
Db	145	ELWYTGSYTEAOWLADLRFVANRYQNPVYVGLDLKNEPHGATGTGNAATDMWKAAR	204
QY	185	AGNAVLNVNPNLLIFVEGVQ-----SYNGDSYVWGGNLQAGQYPVVLNVP-NRLVYSAH	238
Db	205	GSAAVLAVAPKWLIAVEGITDNPVCSTNG-GIFWGGNLQPLACTP--LNIPANRLLAPH	261
QY	239	DIATSYVPTWFSDDPTFPNNPVGIVNKNWGYLFNQNIAPVWLGEFGTTL--QSTTDQTL	296
Db	262	VYGPDVVFSYSDNSFNPNPALWERHFGQFAGTH--ALLGEFGKYGEGDARDKTWQ	319
QY	297	KTLVQVIREPTAQYAGDSFQMTFWSNPDSDGTGILKDDQWT	339
Db	320	DALVKYLR--SKGIN--QGFVWSNPNPSGDTGGILRDDWTSV	357
RESULT 5			
US-10-369-493-15740			
; Sequence 15740, Application US/10369493			
; Publication No. US20030233675A1			
; GENERAL INFORMATION:			
; APPLICANT: Cao, Yongwei			
; APPLICANT: Hinkle, Gregory J.			
; APPLICANT: Slater, Steven C.			
; APPLICANT: Goldman, Barry S.			
; APPLICANT: Chen, Xianfeng			
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION			
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES			
; FILE REFERENCE: 38-10(52052)B			
; CURRENT APPLICATION NUMBER: US/10/369,493			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US 60/360,039			
; PRIOR FILING DATE: 2002-02-21			
; NUMBER OF SEQ ID NOS: 47374			
; SEQ ID NO 15740			
; LENGTH: 501			
; TYPE: PRT			
; ORGANISM: Xanthomonas campestris			
; FEATURE:			
; NAME/KEY: unsure			
; LOCATION: (1)..(501)			
; OTHER INFORMATION: unsure at all xaa locations			
US-10-369-493-15740			





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:34:32 ; Search time 41.3333 Seconds

(without alignments)  
2716.899 Million cell updates/sec

Title: US-09-997-504A-12

Perfect score: 1982

Sequence: 1 AGGGYWHTSGREILDANNVP.....VDTKVDGYLPIKSSIFDRV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 segs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*

2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*

3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*

4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*

5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*

6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*

7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*

8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*

9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*

10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*

11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*

12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*

13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*

14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*

15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*

16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*

17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*

18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1974	99.6	562	9	US-09-981-900B-5
2	1036	52.3	397	15	Sequence 5, Appli
3	792.5	40.0	493	15	Sequence 212, App
4	785.5	39.6	483	15	Sequence 16379, A
5	785.5	39.6	501	15	Sequence 16123, A
6	785.5	39.6	518	15	Sequence 15740, A
7	785.5	39.6	535	15	Sequence 15373, A
8	758	38.2	553	9	Sequence 50, Appl
9	751	37.9	458	14	Sequence 2, Appli
10	751	37.9	458	15	Sequence 2, Appli
11	751	37.9	458	15	Sequence 1281, Ap
12	694.5	35.0	514	15	Sequence 20347, A
13	683.5	34.5	592	15	Sequence 21618, A
14	656.5	33.1	369	15	Sequence 17534, A
15	651.5	32.9	421	15	Sequence 9473, Ap
					Sequence 9262, Ap

16	574	29.0	472	15	US-10-369-493-8139	Sequence 8139, Ap
17	455.5	23.0	362	15	US-10-369-493-3080	Sequence 3080, Ap
18	281.5	14.2	353	15	US-10-369-493-15733	Sequence 15733, A
19	281.5	14.2	353	15	US-10-369-493-16119	Sequence 16119, A
20	281.5	14.2	582	10	US-09-927-827-48	Sequence 48, Appl
21	249.5	12.6	533	12	US-10-424-599-176232	Sequence 176232, A
22	246	12.4	375	15	US-10-369-493-12597	Sequence 12597, A
23	237	12.0	397	15	US-10-369-493-3152	Sequence 3152, Ap
24	195.5	9.9	438	12	US-10-425-114-66753	Sequence 66753, A
25	194.5	9.8	314	15	US-10-369-493-15726	Sequence 15726, A
26	194.5	9.8	314	15	US-10-369-493-16112	Sequence 16112, A
27	194.5	9.8	337	15	US-10-369-493-15359	Sequence 15359, A
28	186	9.4	644	16	US-10-437-963-139883	Sequence 139883, A
29	182.5	9.2	917	16	US-10-437-963-181484	Sequence 181484, A
30	181.5	9.2	375	10	US-09-917-378-3	Sequence 3, Appli
31	181.5	9.2	375	10	US-09-917-378-6	Sequence 6, Appli
32	181.5	9.2	762	10	US-09-917-378-1	Sequence 1, Appli
33	179	9.0	356	10	US-09-917-378-7	Sequence 7, Appli
34	176.5	8.9	346	10	US-09-917-378-8	Sequence 8, Appli
35	176	8.9	517	12	US-10-107-431-259	Sequence 259, App
36	168.5	8.5	558	16	US-10-437-963-185049	Sequence 185049, A
37	157.5	7.9	533	12	US-10-424-599-150988	Sequence 150988, A
38	154.5	7.8	722	12	US-10-282-122A-51993	Sequence 51993, A
39	152.5	7.7	268	12	US-10-424-599-284543	Sequence 284543, A
40	151.5	7.6	356	15	US-10-369-493-17756	Sequence 17756, A
41	150	7.6	217	16	US-10-437-963-112718	Sequence 112718, A
42	149.5	7.5	503	12	US-10-424-599-150986	Sequence 150986, A
43	146.5	7.4	315	15	US-10-369-493-9177	Sequence 9177, Ap
44	146.5	7.4	518	10	US-09-769-734-56	Sequence 56, Appl
45	144	7.3	362	15	US-10-369-493-12370	Sequence 12370, A

#### ALIGNMENTS

RESULT 1

US-09-981-900B-5

; Sequence 5, Application US/09981900B

; Patent No. US20020138878A1

; GENERAL INFORMATION:

; APPLICANT: Sticklen, Vasomesh B

; APPLICANT: Magbool, Shahina B

; APPLICANT: Dale, Bruce E

; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULOSE TO FERMENTABLE SUGARS

; FILE REFERENCE: MSU 4.1-539

; CURRENT APPLICATION NUMBER: US/09/981,900B

; CURRENT FILING DATE: 2002-03-18

; PRIOR APPLICATION NUMBER: 60/242,408

; PRIOR FILING DATE: 2000-10-20

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 5

; LENGTH: 562

; TYPE: PRT

; ORGANISM: Acidothermus cellulolyticus

US-09-981-900B-5

Query Match 99.6%; Score 1974; DB 9; Length 562;

Best Local Similarity 99.7%; Pred. No. 3.1e-185;

Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMQLQIKSLGNT 60

Db 42 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMQLQIKSLGNT 101

QY 61 IRLPYSDLLKPGTWPNSINFQNMQLQGLTSLQMDKIYAYAGIQLRIILDRHRDC 120

Db 102 IRLPYSDLLKPGTWPNSINFQNMQLQGLTSLQMDKIYAYAGIQLRIILDRHRDC 161

QY 121 SQSALWYTSSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWGCDPSIDWRL 180

Db 162 SQSALWYTSSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWGCDPSIDWRL 221



US-08-525-697-2  
 ; Sequence 2, Application US/08525697  
 ; Patent No. 5795764  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Christgau, Stephan  
 ; APPLICANT: Andersen, Lene N  
 ; APPLICANT: Kauppinen, Sakari  
 ; APPLICANT: Helldt-Hansen, Hans P  
 ; APPLICANT: Dalboege, Henrik  
 ; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: No. 57957640 No. 5795764disk of No. 5795764th America, Inc.  
 ; STREET: 405 Lexington Avenue, 64th Floor  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: United States of America  
 ; ZIP: 10174-6401  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/525,697  
 ; FILING DATE: 21-SEP-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Harrington, James J.  
 ; REGISTRATION NUMBER: 38,711  
 ; REFERENCE/DOCKET NUMBER: 4004.204-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-867-0123  
 ; TELEFAX: 212-878-9655  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: DK 0486/93  
 ; FILING DATE: 30-APR-1993  
 ; CLASSIFICATION: 435  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 377 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; US-08-525-697-2

Query Match 9.0%; Score 178.5; DB 1; Length 377;  
 Best Local Similarity 23.4%; Pred. No. 7.4e-09;  
 Matches 83; Conservative 56; Mismatches 115; Indels 101; Gaps 21;

Qy 24 AGIN-WFGPCTCYVHGLSRDYRSMQDKISLGYNTIRLPYSDDILKPGTWPNSINF 81  
 Db 50 AGTNSYWGFLTNDDVDLVMSQLASDLKILRWGFNDVNTKPTD-----GTV-----W 99

Qy 82 YQXNQD-----LQGLTSLOWMDKIVAGQIGLRIILD--RHRPDCSGQSAL----- 126  
 Db 100 YQLHANGTSTINTGADGLRLDYVVTGAEGYGVKLLINFNEWTDYGGMOAYTAYGAAA 159

Qy 127 ---WYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACMGCGDPSIDWRLAAE 183  
 Db 160 QTDFTYNTAIQAAYKNYIKAVWSRYSSAAIFAWELANEPR---CQGC-DTSVLYNNWISD 215

Qy 184 RAGNAVLNVNPNLLIFVEGVQSYNGDSYWGNNLQAGQYPPV-----LNVPNELVYSAHD 239  
 Db 216 -TSKYIKLSKSHLVITI-----GDE-GRGLVDSDSGSYPTTYGEGNLTKNLIGISTID 266

Qy 240 YAT-SVYPTQWFSDDPTFPNNMPGWNKNGVLFNQIA-----PVWLGEFGTT----- 286  
 Db 267 FGLHLIYDPSGKTSYD-----WGNW---ITAHAAACKAVGKPCLLBEYGVTSNHCA 315

Qy 287 LQSTTDT-----WLKTLVOYLRPTAQYAGDSFQWTF-NSWNPDSGDT 328  
 Db 316 VESPWQQTAGNATGISGDLWY-----QYGT-----TFSWQSGSPNDGNT 353

RESULT 13  
 US-09-797-464A-4  
 ; Sequence 4, Application US/09797464A  
 ; Patent No. 6630340  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wiltling, Reinhard  
 ; APPLICANT: Bjornvad, Mads Eskelund  
 ; APPLICANT: Kauppinen, Markus Sakari  
 ; APPLICANT: Schulein, Martin  
 ; TITLE OF INVENTION: Family 5 xyloglucanases  
 ; FILE REFERENCE: 6073.200-US  
 ; CURRENT APPLICATION NUMBER: US/09/797,464A  
 ; CURRENT FILING DATE: 2002-02-19  
 ; NUMBER OF SEQ ID NOS: 25  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 4  
 ; LENGTH: 400  
 ; TYPE: PRT  
 ; ORGANISM: Paenibacillus sp.  
 ; US-09-797-464A-4

Query Match 7.1%; Score 141; DB 4; Length 400;  
 Best Local Similarity 22.5%; Pred. NO. 3.3e-05;  
 Matches 86; Conservative 53; Mismatches 131; Indels 112; Gaps 23;

Qy 3 GGYWHTSGREILDANNVPVRIAGINWFGFETCTCYVHGLSRDYRSMQDKISLGYNTIR 62  
 Db 48 GAGWNLGNQLEATVNGVSETA---W-----GNPVW-----TPELIKKVKAAAGFTIR 92

Qy 63 LP--YSDDLKPGTMPN-SINFYQMNQDLOGLSLQVMDKIVAYAGQIGLRIILDRHPD 119  
 Db 93 IPVSYLNHI---GSAPNVTINAAWLNR-----VQTVVDYAYNEGLVWINHGDG 139

Qy 120 CSGQSALWYTSVSEATWISD-----LQALAQRYKGNPTVVGFDLHNEPHDPACMGCGDP 174  
 Db 140 YNSIPGGWLLVNGSNQAAIKKQYKQVQOQIATKFSNNYNERLIFESMNEVFD---GNVGNP 196

Qy 175 SIDWRLAAERAGNAVLNVNPNLL--IFVEGVQSY--NGDSYW-----WGNLQ-CAGQYP 224  
 Db 197 -----NAAYYANLNAVYQIFVDVTRQGTGNNNARWLLIPGWNNTIDYTVNGY 244

Qy 225 VVL-----NVPN---RLVYSAHDY-----ATSVYV---QTFWSDP 253  
 Db 245 FALPTDHFPSGAIPSSQKRIMISAHYYPWDFAGEENGITQWGAATNPSKSTWQED 304

Qy 254 TFPNNMPGIMNK--NWGYLFNQNIAPVWLGEFGT---TLQSTTDTOTWLKTLVQYLRPTA- 307  
 Db 305 YLNAQFKSYDKFVTOGY-----PVVIGFSGSIDKTAYDSTNNVYRQAYAKAVTATAK 357

Qy 308 QYGADSPQWTFWSMNPDSGDTG 329  
 Db 358 KYGAVPVYV-----DNHGNG 372

RESULT 14  
 US-09-797-464A-7  
 ; Sequence 7, Application US/09797464A  
 ; Patent No. 6630340  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wiltling, Reinhard  
 ; APPLICANT: Bjornvad, Mads Eskelund  
 ; APPLICANT: Kauppinen, Markus Sakari  
 ; APPLICANT: Schulein, Martin  
 ; TITLE OF INVENTION: Family 5 xyloglucanases  
 ; FILE REFERENCE: 6073.200-US  
 ; CURRENT APPLICATION NUMBER: US/09/797,464A  
 ; CURRENT FILING DATE: 2002-02-19  
 ; NUMBER OF SEQ ID NOS: 25  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 7  
 ; LENGTH: 348  
 ; TYPE: PRT

QY 338 TVDTVKGYLAP 349  
Db 562 TWDEQKYNFLKP 573

## RESULT 10

US-09-136-574A-43

; Sequence 43, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

; Anderson, Paige

; Gibbs, Moreland

; Bergquist, Peter

; Daniels, Roy

; Morgan, Hugh W.

; Williams, Diane P.

; TITLE OF INVENTION: Compositions and Methods for

; Treating Cellulose Containing Fabrics Using Truncated

; Cellulase Enzyme Compositions

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howson and Howson

; STREET: Spring House Corporate Center, P.O. Box 457

; CITY: Spring House

; STATE: PA

; COUNTRY: USA

; ZIP: 19477

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/136,574A

; FILING DATE: 19-Aug-1998

; CLASSIFICATION: &lt;Unknown&gt;

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/932,571

; FILING DATE: September 19, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Bak, Mary E.

; REGISTRATION NUMBER: 31,215

; REFERENCE/DOCKET NUMBER: 1997US001/CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-540-9200

; TELEFAX: 215-540-5818

; TELEX: &lt;Unknown&gt;

; INFORMATION FOR SEQ ID NO: 43:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1426 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: No. 6294366e

; SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43

Query Match 31.1%; Score 617; DB 3; Length 1426;

Best Local Similarity 35.5%; Pred. No. 3e-50;

Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY 10 GREILDANNPVRIAGINWFGFTCNVYVHGLSRDYSRMLDQIKSLGYNTIRLPYSDDI 69

Db 1022 GNKIVDKGFVMTLTVGNWFGFTGTVNFDGVWSNKLALAIANKGFNLLRVPISAEI 1081

QY 70 L---KPTMKN-SINFQMNQDLGLTSLQVMDKIYAYAGQIGLRILDEH--RPPCSQ 123

Db 1082 ILNWSKGIYKPKININY-VNPELEGLTSLVDFVVKTEVGLKIMLDIHSAKTDAMGH 1140

QY 124 -SALWYTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDP-----ACWGGCDEPS 175

Db 1141 IYPWYTDITPDEYKACEWITERYKNDTIVAFDLKNEPHGKPNQDSVFAKWDNSTDI 1200

QY 176 IDWRLAARAGNAVLNVNPNLLIFVEGVSYNGD-----SYWVGNGNLOGAGQ 222  
Db 1201 NNWKYAAETCAKRLAKPNWMLIVISGIEAYPKDDVTWTSKSSSDYVSTWVGNGLRGVKK 1260  
QY 223 YPVVL-NVPMRLVYSAHDYATSVYPTW-----FSDTFFNNMPEGIWNKNWGYLFNQNIAP 277  
Db 1261 YPINLGOYQNKVYSPHDYGLVYQOPFWFPGFTKDTLYND---CWRDNWNTYIMDNGIAP 1317  
QY 278 VMLGEFGITLQSTTDDOTWLTQVYLRPTAQYGADSFQMTFWSWNPDSGDTGGILKDDWQ 337  
Db 1318 LLIGEWGGLDGGDNKKNWYLRDYII-----ENHIHTFWCYNANSNGDTGSLGVGYDFS 1371  
QY 338 TVDTVKGYLAP 349  
Db 1372 TWDEQKYNFLKP 1383

## RESULT 11

US-08-276-213-2

; Sequence 21, Application US/08276213

; Patent No. 5536655

; GENERAL INFORMATION:

; APPLICANT: Thomas, Steven

; APPLICANT: Layton, Robert

; APPLICANT: Himmel, Michael

; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: National Renewable Energy Laboratory

; STREET: 1617 Cole Boulevard

; CITY: Golden

; STATE: CO

; COUNTRY: USA

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/276,213

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: O'Connor, Edna

; REGISTRATION NUMBER: 29,252

; REFERENCE/DOCKET NUMBER: NREL IR# 94-08

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303)231-1000

; TELEFAX: (303)231-1098

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 38 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

US-08-276-213-2

Query Match

Best Local Similarity 10.3%; Score 205; DB 1; Length 38;

Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYWHYTSGREILDANNPVRIAGINWFGFTCNVYV 38

Db 1 AGGGYWHYTSGREILDANNPVRIAGINWFGFTCNVYV 38

## RESULT 12

Db 199 TDWNLAAERIGKATLKVAPHWLIPEVGTQFTNPKTSSYKMGYNAMWGGNLAIVKDPV- 257  
QY 227 LNVP-NRLVYSAHDYATSVYPTQWFSPT-FPNMNGIWNKNWGYLFNQNIAFPVWLGEFG 284  
Db 258 -NLPRNKLIVSPHYVGPVNPQYFPGAKGFPDNLPIWTHHFGYVLELGYSVWIGFEG 316  
QY 285 TTL---OSTTDTQWTKLVQYLRPTAQYGADSFQWTFWSPNPSGDTGGILKDDWQTV 339  
Db 317 GKYCHGGDPDVIMQNKLVDMW--IENKFCDF--YWSWNPDSGDTGGILQDDWTI 369

## RESULT 8

US-09-869-197-5

; Sequence 5, Application US/09869197

; Patent No. 6566113

; GENERAL INFORMATION:

; APPLICANT: TAKAYAMA, Masanori

; APPLICANT: UNEDA, Kahoko

; APPLICANT: KOYAMA, Kiyozo

; APPLICANT: ASADA, Kiyozo

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: POLYPEPTIDES

; FILE REFERENCE: TAKAYAMA=6

; CURRENT APPLICATION NUMBER: US/09/869,197

; CURRENT FILING DATE: 2001-06-25

; PRIOR APPLICATION NUMBER: PCT/JF99/07009

; PRIOR FILING DATE: 1999-12-14

; PRIOR APPLICATION NUMBER: JP 366237/1998

; PRIOR FILING DATE: 1998-12-24

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 5

; LENGTH: 458

; TYPE: PRT

; ORGANISM: Pyrococcus horikoshii OT3

US-09-869-197-5

Query Match 37.9%; Score 751; DB 4; Length 458;  
Best Local Similarity 45.0%; Pred. No. 6.8e-64;  
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGEILDANNVPVRIAGINWFGTCTNYVHGLRSRDYRSMLEQIKSLGYNIRLPYS 67  
Db 54 TSGE-----TPHRLGVNWFGETPNVHVLGKWNWEDMLLQIKSLGFNAIRLPFCT 107  
QY 68 DILKPGWPNINSIFYQMNQDQLGTSIQVMDKIVAYAGQIGRLIILDRHPRDCSGSALW 127  
Db 108 ESVKPGTQIGID-YSKNPDJRGDLS-QIMEKIIKAGDGLIFVLLDYHHRIGCTHIEPLW 166  
QY 128 YTSVSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWGCGDPS 175  
Db 167 YTEFSEDFNTWIEVAKRGKWNWIGADLKNHPSVTSPPAAVTDGTGATWGNPA 226  
QY 176 IDWLAERAGNAVLSPNPNLLIFVEGVQSVN-----GDSYWWGGNLQAGQYPPV 226  
Db 227 TDWNLAAERIGKATLKVAPHWLIPEVGTQFTNPKTSSYKMGYNAMWGGNLAIVKDPV- 285  
QY 227 LNVP-NRLVYSAHDYATSVYPTQWFSPT-FPNMNGIWNKNWGYLFNQNIAFPVWLGEFG 284  
Db 286 -NLPRNKLIVSPHYVGPVNPQYFPGAKGFPDNLPIWTHHFGYVLELGYSVWIGFEG 344  
QY 285 TTL---OSTTDTQWTKLVQYLRPTAQYGADSFQWTFWSPNPSGDTGGILKDDWQTV 339  
Db 345 GKYCHGGDPDVIMQNKLVDMW--IENKFCDF--YWSWNPDSGDTGGILQDDWTI 397

## RESULT 9

US-09-136-574A-47

; Sequence 47, Application US/09136574A

; Patent No. 6294366

; GENERAL INFORMATION:

; APPLICANT: Farrington, Graham K.

Anderson, Paige  
Gibbs, Moreland  
Bergquist, Peter  
Daniels, Roy  
Morgan, Hugh W.  
Williams, Diane P.  
TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing Fabrics Using Truncated  
Cellulase Enzyme Compositions  
NUMBER OF SEQUENCES: 49  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997TUS001/CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 616 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 47:  
US-09-136-574A-47

Query Match 31.1%; Score 617; DB 3; Length 616;  
Best Local Similarity 35.5%; Pred. No. 8.6e-51;  
Matches 132; Conservative 70; Mismatches 126; Indels 42; Gaps 11;  
QY 10 GREILDANNVPVRIAGINWFGTCTNYVHGLRSRDYRSMLEQIKSLGYNIRLPYSDDI 69  
Db 212 GNKIVDKGKPVMLGVNWFGETCTNVFDGVWSCNLKSALAEIANRGFNLLRVPISAEI 271  
QY 70 L---KPGTNP--SIFYQMNQDQLGTSIQVMDKIVAYAGQIGRLIILDRH--RDCSGQ 123  
Db 272 ILNWSKGYPKPNINYY-VNPELGLTSLVDFVYVKTCKEVLKIMLDIISAKTDAMGH 330  
QY 124 -SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----ACWGCGDPS 175  
Db 331 IYPVAYTOTITPEDYKACEWITERYKNDTIVAFDLKNEPHGKWPQSVFAKWDNSTDI 390  
QY 176 IDWLAERAGNAVLSPNPNLLIFVEGVQSVNCD-----SYWWGGNLQAGQ 222  
Db 391 NNWKYAAETCAKRIIAKPNMLIVIEGIEAVPKDDVTWTSKSSDYSTWGGNLRGVKK 450  
QY 223 YPVVL-NVPNRLVYSAHDYATSVYPTW---FSDPTFPNNMNGIWNKNWGYLFNQNIAFP 277  
Db 451 YPINLGOYKNVYSPHYDYGPLVYQPPYFPGFTKDTLYND---CWRDNWYIMDNGIAP 507  
QY 278 VWLGEFTTLQSTTQWTKLVQYLRPTAQYGADSFQWTFWSPNPSGDTGGILKDDWQ 337  
Db 508 LLIGWGGYLDGGDNKRWMLRDYII-----ENHIHTFWCYNANSNGDTGGILVGYDFS 561

Query Match 38.2%; Score 758; DB 3; Length 553;  
Best Local Similarity 45.2%; Pred. NO. 1.9e-64;  
Matches 160; Conservative 50; Mismatches 108; Indels 36; Gaps 11;

QY 13 ILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYSMLDQIKSLGYNTRLRPLYSDDLKPK 72  
DB 4 VATGEETPIHLFGVNNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLFFCTQSVKP 63

QY 73 GTPNSINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQSALWYTSV 132  
DB 64 GTMPTAID-YAKNPDLQGLDSVQIMEKIIKAGDLGIFVLLDYHRIICNFIEPLWYTSDF 122

QY 133 SEATWISDLQALQRYKGNPTVVGFDLHNEPDP-----ACWCGDPSIDWRL 180  
DB 123 SEQDYINTWVEVAQRFGKYWNVI GADLKNP HSSSPAPAAAYTGGSGATWGMGNATDNL 182

QY 181 AERAGNAVLNPNLLIFVEGVQ-----SYN-GDSYMWGMLQAGQYPPVVLNVP- 230  
DB 183 AERIGRAILEVAPOWIFVEGTQFTTPEIDGRYKWHNAWGNLMGVKYPV--NLPR 240

QY 231 NRVYSADYATSVYPTWFSDP--TFPNNMFGIWNKNGYLFNQNIAPIVWLGEFGTTL- 287  
DB 241 DKVYSPQVYGVSEYDQPYF-DPEGFPDNLPEIWHYHFGYVKLDLGYVWVIGFEGGYG 299

QY 288 --QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQTV 339  
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDFP---YWSWNPNSGDTGGILKDDWTTI 348

RESULT 6  
US-09-430-669-2  
; Sequence 2, Application US/09430669  
; Patent No. 6329187  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; FILING DATE: 28-Oct-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,572  
; FILING DATE: 22-MAY-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Hezron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS: <unknown>  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-430-669-2

Query Match 38.2%; Score 758; DB 4; Length 553;  
Best Local Similarity 45.2%; Pred. NO. 1.9e-64;  
Matches 160; Conservative 50; Mismatches 108; Indels 36; Gaps 11;

QY 13 ILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYSMLDQIKSLGYNTRLRPLYSDDLKPK 72  
DB 4 VATGEETPIHLFGVNNWFGFETPNVYVHGLWSRWEDMLLQIKSLGFNAIRLFFCTQSVKP 63

QY 73 GTPNSINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQSALWYTSV 132  
DB 64 GTMPTAID-YAKNPDLQGLDSVQIMEKIIKAGDLGIFVLLDYHRIICNFIEPLWYTSDF 122

QY 133 SEATWISDLQALQRYKGNPTVVGFDLHNEPDP-----ACWCGDPSIDWRL 180  
DB 123 SEQDYINTWVEVAQRFGKYWNVI GADLKNP HSSSPAPAAAYTGGSGATWGMGNATDNL 182

QY 181 AERAGNAVLNPNLLIFVEGVQ-----SYN-GDSYMWGMLQAGQYPPVVLNVP- 230  
DB 183 AERIGRAILEVAPOWIFVEGTQFTTPEIDGRYKWHNAWGNLMGVKYPV--NLPR 240

QY 231 NRVYSADYATSVYPTWFSDP--TFPNNMFGIWNKNGYLFNQNIAPIVWLGEFGTTL- 287  
DB 241 DKVYSPQVYGVSEYDQPYF-DPEGFPDNLPEIWHYHFGYVKLDLGYVWVIGFEGGYG 299

QY 288 --QSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQTV 339  
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDFP---YWSWNPNSGDTGGILKDDWTTI 348

RESULT 7  
US-09-869-197-1  
; Sequence 1, Application US/09869197  
; Patent No. 6566113  
; GENERAL INFORMATION:  
; APPLICANT: TAKAYAMA, Masanori  
; APPLICANT: UMEDA, Kanoko  
; APPLICANT: KOYAMA, No. 6566113uto  
; APPLICANT: ASADA, Kiyozi  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: TAKAYAMA-6  
; CURRENT APPLICATION NUMBER: US/09/869,197  
; CURRENT FILING DATE: 2001-06-25  
; PRIOR APPLICATION NUMBER: PCT/JF99/07009  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: JP 366237/1998  
; PRIOR FILING DATE: 1998-12-24  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Pyrococcus horikoshii OT3  
US-09-869-197-1

Query Match 37.9%; Score 751; DB 4; Length 430;  
Best Local Similarity 45.0%; Pred. No. 6.2e-64;  
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYSMLDQIKSLGYNTRLRPLYS 67  
DB 26 TSGE-----TPHLFGVNNWFGFETPNVYVHGLWKNWEDMLLQIKSLGFNAIRLFPCT 79

QY 68 DILKPGTNPNSINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQSALW 127  
DB 80 ESKVFGTQPIGID-YSKNPDLRGLDSLQIMEKIIKAGDLGIFVLLDYHRIICGCTHIBPLW 138

QY 128 YTSVVSATWISDLQALQRYKGNPTVVGFDLHNEPDP-----PACWCGDPS 175  
DB 139 YTEDFSEEDFINTWIEVAKRFGKYWNVI GADLKNP HSSVPAAAYTGGTATWGMGNA 198

QY 176 IDWRLAERAGNAVLNPNLLIFVEGVQSYN-----GDSYMWGMLQAGQYPPV 226





Query Match 99.6%; Score 1974; DB 1; Length 358;  
Best Local Similarity 99.7%; Pred. No. 6.7e-182;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYRSMIDQIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYRSMIDQIKSLGYNT 60

QY 61 IRLPYSDDDILKPGTWPNSINFYQMNQDQLGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120  
DB 61 IRLPYSDDDILKPGTWPNSINFYQMNQDQLGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120

QY 121 SGOSALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRL 180  
DB 121 SGOSALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRL 180

QY 181 AERAGNAVLSVNPILLIFVEGVQSYNGDSYMWGGLQAGQYPVVLNPNRLVYSAHDY 240  
DB 181 AERAGNAVLSVNPILLIFVEGVQSYNGDSYMWGGLQAGQYPVVLNPNRLVYSAHDY 240

QY 241 ATSVYPTWFSDFTPFNNMFGIWNKNWGYLFNQNIAPVWLGEFGTTLOSTTDQTLWLTIV 300  
DB 241 ATSVYPTWFSDFTPFNNMFGIWNKNWGYLFNQNIAPVWLGEFGTTLOSTTDQTLWLTIV 300

QY 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDQWQVTDVTKDGYLAPIKSSIFDPV 358  
DB 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDQWQVTDVTKDGYLAPIKSSIFDPV 358

RESULT 2  
US-08-276-213-3  
; Sequence 3, Application US/08276213  
; Patent No. 5536655  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Steven  
; APPLICANT: Laymon, Robert  
; APPLICANT: Himmel, Michael  
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: National Renewable Energy Laboratory  
; STREET: 1617 Cole Boulevard  
; CITY: Golden  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80401-3393  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/276,213  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: O'Connor, Edna  
; REGISTRATION NUMBER: 29,252  
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (303)231-1000  
; TELEFAX: (303)231-1098  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 521 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-08-276-213-3

Query Match 99.6%; Score 1974; DB 1; Length 521;  
Best Local Similarity 99.7%; Pred. No. 1.2e-181;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYRSMIDQIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDRYRSMIDQIKSLGYNT 60

QY 61 IRLPYSDDDILKPGTWPNSINFYQMNQDQLGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120  
DB 61 IRLPYSDDDILKPGTWPNSINFYQMNQDQLGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120

QY 121 SGOSALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRL 180  
DB 121 SGOSALWYTSVSEATWISDLQALQARYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRL 180

QY 181 AERAGNAVLSVNPILLIFVEGVQSYNGDSYMWGGLQAGQYPVVLNPNRLVYSAHDY 240  
DB 181 AERAGNAVLSVNPILLIFVEGVQSYNGDSYMWGGLQAGQYPVVLNPNRLVYSAHDY 240

QY 241 ATSVYPTWFSDFTPFNNMFGIWNKNWGYLFNQNIAPVWLGEFGTTLOSTTDQTLWLTIV 300  
DB 241 ATSVYPTWFSDFTPFNNMFGIWNKNWGYLFNQNIAPVWLGEFGTTLOSTTDQTLWLTIV 300

QY 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDQWQVTDVTKDGYLAPIKSSIFDPV 358  
DB 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGLKDDQWQVTDVTKDGYLAPIKSSIFDPV 358

RESULT 3  
US-08-651-572-2  
; Sequence 2, Application US/08651572  
; Patent No. 5789228  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILLILLAN,  
; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,572  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Herron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
US-08-651-572-2

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:31:56 ; Search time 14.6667 Seconds  
(without alignments)  
1260.143 Million cell updates/sec

Title: US-09-997-504A-12  
Perfect score: 1982  
Sequence: 1 AGGGYWHTSREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues  
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pap.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pap.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pap.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pap.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pap.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1974	99.6	358	1	US-08-604-913B-11
2	1974	99.6	521	1	US-08-276-213-3
3	758	38.2	553	1	US-08-651-572-2
4	758	38.2	553	3	US-09-066-544-2
5	758	38.2	553	3	US-08-951-086-2
6	758	38.2	553	4	US-09-430-669-2
7	751	37.9	430	4	US-09-869-197-1
8	751	37.9	458	4	US-09-869-197-5
9	617	31.1	616	3	US-09-136-574A-47
10	617	31.1	1426	3	US-09-136-574A-43
11	205	10.3	38	1	US-08-276-213-2
12	178.5	9.0	377	1	US-08-555-697-2
13	141	7.1	400	4	US-09-797-464A-4
14	137	6.9	348	4	US-09-797-464A-7
15	136	6.9	395	4	US-09-797-464A-2
16	134	6.8	363	4	US-09-797-464A-11
17	133.5	6.7	663	4	US-09-134-078-61
18	133.5	6.7	680	4	US-09-134-078-25
19	132	6.7	476	4	US-09-338-159B-4
20	132	6.7	490	4	US-09-338-159B-2
21	131	6.6	24	1	US-08-276-213-1
22	127	6.4	331	4	US-09-338-159B-12
23	126	6.4	468	4	US-09-485-648-4
24	126	6.4	468	4	US-09-503-565-4
25	126	6.4	468	4	US-09-485-649-4
26	126	6.4	468	4	US-09-338-159B-8
27	126	6.4	493	4	US-09-485-648-2

28	126	6.4	493	4	US-09-503-565-2
29	126	6.4	493	4	US-09-485-649-2
30	126	6.4	493	4	US-09-338-159B-6
31	124	6.3	320	4	US-09-339-159B-22
32	114	5.8	360	4	US-09-134-078-27
33	113	5.7	461	1	US-08-672-571A-3
34	113	5.7	490	1	US-08-672-571A-1
35	112.5	5.7	429	1	US-08-745-977-4
36	112.5	5.7	429	3	US-09-040-699A-4
37	111	5.6	327	2	US-08-169-948B-16
38	111	5.6	327	2	US-08-448-873-16
39	111	5.6	327	3	US-08-382-452D-16
40	111	5.6	327	4	US-08-507-362A-8
41	111	5.6	327	4	US-09-916-494A-16
42	111	5.6	418	3	US-09-254-733-5
43	109.5	5.5	430	2	US-08-924-440-2
44	108	5.4	317	2	US-09-066-075-2
45	108	5.4	317	2	US-08-518-615A-2

ALIGNMENTS

RESULT 1  
US-08-604-913B-11  
; Sequence 11, Application US/08604913B  
; Patent No. 5712142  
; GENERAL INFORMATION:  
; APPLICANT: Adney, William S.  
; APPLICANT: Thomas, Steven R.  
; APPLICANT: Himmel, Michael E.  
; APPLICANT: Baker, John O.  
; APPLICANT: Chou, Yat-Chen  
; TITLE OF INVENTION: METHOD FOR INCREASING  
; NUMBER OF SEQUENCES: 12  
; TITLE OF INVENTION: THERMOSTABILITY IN CELLULASE ENZYMES  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: National Renewable Energy Laboratory  
; STREET: 1617 Cole Boulevard  
; CITY: Golden  
; STATE: CO  
; COUNTRY: U.S.A.  
; ZIP: 80401-3393  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASC II (DOS) text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/604,913B  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/276,213  
; FILING DATE: 15-07-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Edna M. O'Connor  
; REGISTRATION NUMBER: 29,252  
; REFERENCE/DOCKET NUMBER: 95-56  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/384-7499  
; TELEFAX: 303/384-7499  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 358 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; FEATURE:  
; NAME/KEY: E1-CAT  
; US-08-604-913B-11

PI Anderson P, Bergquist PL, Daniels RM, Farrington GK, Gibbs MD;  
PI Morgan H, Williams DP;  
XX  
XX WPI; 1999-315403/27.  
DR N-PSDB; AAX55660.  
XX  
XX  
PT New truncated cellulase proteins, useful in detergents and for producing  
PT 'stonewashed' denim.  
XX  
XX Claim 7; Page 42-43; 65pp; English.  
XX  
XX The invention relates to a recombinant cellulase active protein free of  
CC proteinases of native thermophilic and alkaliphilic origin, comprising  
CC the truncated sequences Cel B5, Cel B4/5, Cel B1, Cel E1/2, Cel 1/2/3,  
CC Cel 6 or Cel E3/B5, or a stability region from one of the defined full-  
CC length sequences, or functional equivalents. Cel B5 extends from amino  
CC acid A1011 to P1424 or N1425 or N1426, and Cel B4/5 extends from amino  
CC acid K635 to N1426 in the sequence shown in AAY13492; Cel E1 extends from  
CC amino acid Y39 to D481, Cel E1/2 extends from Y39 to G635, Cel E1/2/3  
CC extends from Y39 to G812, Cel E6 extends from amino acid V1233 to K1751  
CC and the stability region extends from amino acid E482 to G635 in the  
CC sequence shown in AAY13493; Cel E3/B5 is shown in AAY13494. The new  
CC enzymes are useful in laundry detergent compositions to prevent or remove  
CC staining, backstaining or graying, for use on cellulosic materials  
CC including cotton-containing fabrics. They are especially useful for  
CC preventing redeposition of colorant during stonewashing, and for  
CC processing of textiles where cellulose breakdown is required. The new  
CC truncated enzymes show reduced redeposition of dye compared to using non-  
CC truncated cellulase compositions  
XX  
XX Sequence 616 AA;  
SQ  
Query Match 31.1%; Score 617; DB 2; Length 616;  
Best Local Similarity 35.5%; Pred. No. 1.2e-49;  
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;  
QY 10 GREILDANNVPVRIAGINFGFTCNVYVHGLRSRDYRSMLOQIKSLGYNITRLPYSDDI 69  
Db 212 GKKIVDKGKPKVWLTGVNMFNTGTNFDGVWSCNLSKALAEIANRGNLLEVPISAEI 271  
QY 70 L---KPTMPN-SINFQNMDD:QGLTSLQVMDKI VAVAGQIGRLIILDRH--RPDCSQ 123  
Db 272 ILNWSKGIYKPKNINYY-VNPELEGLTSLVEFDVFKTCVEGLKIMLDIHSAKTDAMGH 330  
QY 124 -SALWYTSVSVEATWISDIQALAQRYKGNPTVVYVGLHNEPHDP-----ACWGGDPS 175  
Db 331 IYPVVVTDTITPDYKACAEWITERYKNDDTIVAFDLKNEPRGKQWQDSVFAKWDNSTDI 390  
QY 176 IDWRLAERAGNAVLSNPNLLIFVGVQSYNGD-----SYWVGSLQGAGQ 222  
Db 391 NNWYAAETCAKRLAKNPNMLIVIEGIEAPKDDVTWTSKSSDYSTWVGSLNGVKK 450  
QY 223 YPVVL-NVPNRLVYSAHDYATSYVPQW----FSDPTFPNMPGIWKNWGYLFNQNIAP 277  
Db 451 YPINLGQYQNKVYVSPHDYGLVYQQPWFYFGFTKDTLYND---CWRDNWYIMDNGIAP 507  
QY 278 VWLGEFTTLQSTDDTQTLKTLVQYLRPTAQYGADSFQWTSWNPDSGDTGILKDDWQ 337  
Db 508 LLIGEWGGYLDGDNKWKMYLDRYII-----ENHIHTFCYNANSNGDTGLGVGYDFS 561  
QY 338 TVDTVKDGYLAP 349  
Db 562 TWDEQKYNFLKP 573  
Search completed: August 2, 2004, 16:34:25  
Job time : 49.6667 secs

PT Polypeptides with heat-resistant cellobiohydrolase activity for efficient  
 XX breakdown of cellulose biomass.  
 PS Disclosure; Page 45-47; 50pp; Japanese.

XX This invention describes a novel polypeptide originating in *Pyrococcus*  
 CC horikoshii O3 which has cellobiohydrolase activity. The polypeptide of  
 CC the invention is capable of decomposing poly(D-glucopyranose) having beta  
 CC -1,4 bonds and can be used for the efficient and straightforward  
 CC breakdown of cellulose biomass to glucose. This sequence represents a P.  
 CC horikoshii O3 cellobiohydrolase associated protein described in the  
 CC method of the invention

XX Sequence 458 AA;

Query Match 37.9%; Score 751; DB 3; Length 458;  
 Best Local Similarity 45.0%; Pred. No. 1.2e-62;  
 Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNTIRLPYS 67  
 Db 54 TSGEE-----TPHLFGVNWFGFETPNHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107  
 Qy 68 DILKPGTWPNSINFYQWNOQLGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGOSALW 127  
 Db 108 ESKVFGTQPID-YSKNPDLRGLDSLQIMEKIIKAGDLGIFVLLDYHRICTHIEPLW 166  
 Qy 128 YTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWGGDPS 175  
 Db 167 YTEDFSDEDFINTWIEVAKRFGKYNVIGADLKNEPHSVTSPPAAYTDGTGATWGMGNA 226  
 Qy 176 IDWRLAERAGNAVLSVNPILLIFVEGVQSTN-----GDSYWGNGNLQAGQYPPV 226  
 Db 227 TDWNLAAERAGKALKVAPHWLIIFVEGTQTNPKTDSYKNGYNWNGNLMAVKQYPPV- 285  
 Qy 227 LNVP-NRLVSAHDYATSVYPTQWFSPT-FPNNMPCGINKNWGLYFNQNIAPVNLGEFG 284  
 Db 286 -NLRNKLIVSPHYGPDVYNQPYFGAKGPDNLPIWYHHFGYKVLGLGYSVWIGFEG 344  
 Qy 285 TTL---QSTDTQWLKTLVQLRPTAQYAGDSFQWTFWNPDSGDTGGILKDDQWTV 339  
 Db 345 GYGHGGDPRDVIWQNLVDMW--IENKFCDF--YWSNPNDSGDTGGILQDDWTII 397

RESULT 14

AAB96264  
 ID AAB96264 standard; protein; 514 AA.

XX AAB96264;

AC AAB96264;

DT 29-OCT-2001 (first entry)

DB Putative P. abyssi endoglucanase #2.

KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX *Pyrococcus abyssi*.

OS FR2792651-A1.

XX 27-OCT-2000.

PF 21-APR-1999; 99FR-00005034.

XX 21-APR-1999; 99FR-00005034.

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX (IFRE-) IFREMER INST FR RECH EXPL MER.

PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;

XX Querellou J, Weissenbach J, Saurin W, Heilig R;

DR WPI; 2001-126236/14.

XX New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins  
 PT useful in industry.

XX Claim 7; Page 921-922; 1657pp; French.

XX The present invention relates to the genomic sequence of *Pyrococcus*  
 CC abyssi (see AAF8431 and AAH41223-7) and P. abyssi proteins. P. abyssi is  
 CC a hyperthermophilic archaeon, which is isolated from deep-sea  
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade. Note: This patent is in the same patent family as  
 CC WO200005062, which contains additional sequences as shown in AAB99132-  
 CC AAB99143, AAH75903-AAH75920 and AAG66436

XX Sequence 514 AA;

Query Match 35.0%; Score 694.5; DB 4; Length 514;  
 Best Local Similarity 41.9%; Pred. No. 3.6e-57;  
 Matches 149; Conservative 57; Mismatches 121; Indels 29; Gaps 10;

Qy 7 HTSGREILDANNV-----PVRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNT 60  
 Db 30 YTAENGIIIFVQNVTTGEKKPLYLHGVSWFGFELKHVVYGLDKRKNWDLKDKVRLGFGNA 89  
 Qy 61 IRLPYSDDLKPGTWPN--SINFYQWNOQLGLTSLQVMDKIVAYAGQIGLRIILDRHRP 118  
 Db 90 IRLPCSSIRPTDPSPERIN-YELNPDKNLSLEIMEXIIYVANSIGLYLLDYHRI 148  
 Qy 119 DCSGOSALWYSSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWGGDPSIDW 178  
 Db 149 GCEETPLWYTENYSEQYIKDWIFLAKRFKGPVNVIGADIKNEPHGEGAGTGDGR-DF 207  
 Qy 179 RLAAERAGNAVLSVNPILLIFVEGVQ-----SYNGDSYV--WGNLQAGQYPPVL 227  
 Db 208 RLFAEKVGRILKVAPHWLIIFVEGTQYTHVNPIDIEIKKGWTFWGNLGVKDYPPVL 267  
 Qy 228 NVNPLVYSAHDYATSVYPTQWFSPTFPNNMPCGINKNWGLYFNQNIAPVNLGEFGTTL 287  
 Db 268 -PRGVVYSPHYGPDVYNQPYFGAKGPDNLPIWYHHFGYKVLGLGYSVWIGFEG 325  
 Qy 286 QSTDTQWLKTLVQLRPTAQYAGDSFQWTFWNPDSGDTGGILKDDQWTV 343  
 Db 326 EG-LDKVWQDAFVKWLKKIY-----NFFYVCLNPESGDTGGIFLDKDKVNWEEK 375

RESULT 15

AAY13494

ID AAY13494 standard; protein; 616 AA.

XX AAY13494;

DT 30-JUL-1999 (first entry)

DE Truncated cellulase Cel E3/B5.

XX Cellulase; proteinase, truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 KW cotton-containing fabric; stonewashing.

OS Unidentified.

XX EP921188-A2.

PD 09-JUN-1999.

XX 15-SEP-1998; 98EP-00810919.

XX 19-SEP-1997; 97US-00932571.

XX (CLRN ) CLARIANT FINANCE BVL LTD.

XX





KW emulsion stability; temperature resistance; pseudoplasticity; amylase;  
KW cellulase; extracellular protease; intracellular protease;  
KW glucose dehydrogenase; enzyme.  
XX  
OS Xanthomonas campestris.  
XX  
PN US2003036176-A1.  
XX  
PD 20-FEB-2003.  
XX  
PF 10-AUG-2001; 2001US-00927827.  
XX  
PR 28-MAR-2001; 2001US-0279493P.  
XX  
PR (BOWE/) BOWER S G.  
PA (RAMS/) RAMSEIER T M.  
XX  
PI Bower SG, Ramseier TM;  
XX WPI; 2003-625389/59.  
DR N-PSDB; ADD24896.  
XX  
XX New transformed cell or organism having reduced or enhanced activity of  
PT at least one protein, useful for producing xanthan gum, which are useful  
PT for providing formulations and properties.  
XX  
XX Claim 1; SEQ ID NO 50; 135bp; English.  
XX  
XX The present invention relates to polypeptide and polynucleotide sequences  
CC from Xanthomonas campestris which may be used for activity reduction or  
CC enhancement using directed genetic engineering. A transformed cell or  
CC organism having reduced or enhanced activity of at least one such protein  
CC e.g. galactomannanase can be generated by disrupting the gene encoding  
CC the protein. The activity of the protein is reduced by the presence of an  
CC antisense nucleic acid sequence. The nucleic acid sequence of the gene  
CC encoding the protein is a recombinant sequence having at least one  
CC mutation as compared to the wild-type gene encoding the protein. The  
CC transgenic cell or microorganism are useful for producing xanthan gum,  
CC which are useful for providing formulations and properties, such as long-  
CC term suspension and emulsion stability in alkaline, acid, and salt  
CC solutions, temperature resistance, and pseudoplasticity. The present  
CC sequence represents an enzyme relating to the present invention.  
XX  
SQ Sequence 535 AA;  
Query Match 39.6%; Score 785.5; DB 7; Length 535;  
Best Local Similarity 46.6%; Pred. No. 7.9e-66;  
Matches 160; Conservative 59; Mismatches 105; Indels 19; Gaps 9;  
QY 5 YHSTGREILDANNVPRIAGINWGFETCNVYVHGLSRDYSRLDQIKSLGYNTIRLP 64  
DB 77 YSINNRSQIVDSGKVVQLKGVNVFETGNGHVMHGLMARNMKDMIVQMGGJGFNAVRLP 136  
QY 65 YSDIILKPGTMPNSINFYQNNQDLQSLTSLQWMDKIYVAGQIGLRIILDRHRPCSGQS 124  
DB 137 FCPATLRSDDTPASID-YSRNADLQSLTSLQWMDKIYVAGQIGLRIILDRHRPCSGQS 195  
QY 125 ALWYTSVSEATWISDLOALQRYKGNPTVGFPLHNEPHDPACWCGDPSIDWRLAER 184  
DB 196 ELWYTSVTEAQLADLRFNARYKVPVYGLDLXNEPHCATGTGNTAATDNWKAER 255  
QY 185 AGNAVLNPNLLIFVEGVQ-----SYNGDSYVWGNLQAGQYFVVLNVP-NRLVYSAH 238  
DB 256 GSAALVAVAPKWLIAVEGITDNPVCSITNG-GIFWGNLQPLACTP--LNIPANRLLALPH 312  
QY 239 DYATSVYPTWESDPTFNNMPCGMKNWGLFNQNIAPFWLGEFTTL--QSTTDOTWL 296  
DB 313 VYGPDFVQSYFNDSPNPNMIPAWEHFQCFAGTH--ALLGEFDGCKYGEGBDARDKWQ 370  
QY 297 KTLVQLRPTAQYGADSFQWTFWSMNPDSGDTGGILKDDWQTV 339  
DB 371 DALVKYLR---SKGIN--QGFYWSWNPNSGDTGILRDDWTSV 408

RESULT 9  
AAW34999  
ID AAW34999 standard; protein; 628 AA.  
XX  
AC AAW34999;  
XX  
XX 27-AUG-2003 (revised)  
DT 21-MAY-1998 (first entry)  
XX  
DE Archaeobacterium AEP11a endoglucanase.  
XX  
XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;  
KW beta-1,4-glycosidic bond; hydrolysis; saccharification;  
KW thermostable enzyme; thermophilic; glycosidase.  
XX  
OS archaeson.  
XX Unidentified..  
XX WO9744361-A1.  
PN 27-NOV-1997.  
XX  
PD 22-MAY-1997;  
XX 97WO-US008793.  
PF 22-MAY-1996;  
XX 96US-00651572.  
PR (RECO-) RECOMBINANT BIOCATALYSIS INC.  
PA  
XX Lam DE, Mathur EJ;  
XX  
XX WPI; 1998-018435/02.  
DR N-PSDB; AAT94207.  
XX  
XX Endo:glucanase(s), preferably form archael bacterium, AEP11 la - useful  
PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic  
PT bonds in cellulose.  
XX  
XX Claim 1; Fig 10; 164pp; English.  
XX  
XX This protein comprises an endoglucanase of archaeobacterium AEP11a (Clone  
CC 630F2), a hydrothermal vent isolate. The endoglucanase is capable of  
CC degrading carboxymethylcellulose and of hydrolysing the beta-1,4-  
CC glycosidic bonds in cellulose. It has homology to another endoglucanase  
CC (see AAW34985) of archaeobacterium AEP11a. It can be produced from native  
CC cells or from recombinant host cells, especially prokaryotic host cells  
CC transformed with a plasmid or virus-derived vector including the  
CC endoglucanase DNA (see AAT94207). 24 Endoglucanases (see AAW34986-W35008)  
CC are claimed. They can be used to degrade cellulose for the conversion of  
CC plant biomass into fuels and chemicals, for use in detergents, textiles,  
CC animal feed, waste treatment, and in the fruit juice and brewing  
CC industries for the clarification and extraction of juices. (Updated on 27  
CC -AUG-2003 to correct OS field.)  
XX  
SQ Sequence 628 AA;  
Query Match 38.3%; Score 759; DB 2; Length 628;  
Best Local Similarity 45.5%; Pred. No. 3.4e-63;  
Matches 161; Conservative 48; Mismatches 109; Indels 36; Gaps 11;  
QY 13 ILDANNVPRIAGINWGFETCNVYVHGLSRDYSRLDQIKSLGYNTIRLPYSDIILK 72  
DB 4 VATGEETPIHLFGVNFVGFETPNVYVHGLSRNWMELMLQIKSLGFNAIRLPCTQSVKP 63  
QY 73 GTMPNSINFYQNNQDLQSLTSLQWMDKIYVAGQIGLRIILDRHRPCSGQSALWYTSV 132  
DB 64 GTMPTAID-YAKNPDQLQSLDSVQIMEXIKKAGDLGIFVLLDYHRCNFIPLWYDTSF 122  
QY 133 SEATWISDLOALQRYKGNPTVGFPLHNEPHDP-----ACWCGDPSIDWRL 180  
DB 123 SEQDYINTWVEVAFQKYNWVIGADLKNPEHSSSPAPAAVYDGSATWGMGNATDWNL 182  
QY 181 AAERAGNAVLNPNLLIFVEGVQ-----SYN-GDSYVWGNLQAGQYFVVLNVP- 230

CC activity comprises replacing an active site-associated amino glycosyl-  
 CC stabilising amino acid with an amino acid that does not strongly bind a  
 CC disaccharide product in the active site. Conversely, the method for  
 CC making a glycosyl hydrolase with increased soluble substrate catalytic  
 CC activity comprises replacing a hydrophobic substrate-binding amino acid  
 CC with a positively charged residue. The invention also discloses mutants  
 CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)  
 CC produced according to the method of the invention. The Y245G mutant  
 CC (AAB48788) has improved activity with insoluble substrates, and the W42R  
 CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with  
 CC soluble substrates. The invention also encompasses DNA encoding these  
 CC mutants. The glycosyl hydrolases of the invention are used as catalysts  
 CC for cellulose hydrolysis to produce sugars that can be fermented to  
 CC produce fuels such as ethanol. The present sequence represents the  
 CC Acidothermus cellulolyticus E1 endoglucanase Y82R mutant  
 XX  
 SQ Sequence 521 AA;

Query Match 99.1%; Score 1965; DB 4; Length 521;  
 Best Local Similarity 99.4%; Pred. No. 1.9e-178;  
 Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGYYHTSGREILDANNVPVRVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60  
 DB 1 AGGYYHTSGREILDANNVPVRVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60  
 QY 61 IRLPYSDILKPGTWPNSINFYQNMQLGLTSLOVMDKIYVAGQIGLRILDRHRPDC 120  
 DB 61 IRLPYSDILKPGTWPNSINFYQNMQLGLTSLOVMDKIYVAGQIGLRILDRHRPDC 120  
 QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 QY 181 AAERAGNAVLSVNPENLLIFVEGVQSYNGDSYWMGGLQAGQYPPVNLNPNRLVYSAHDY 240  
 DB 181 AAERAGNAVLSVNPENLLIFVEGVQSYNGDSYWMGGLQAGQYPPVNLNPNRLVYSAHDY 240  
 QY 241 ATSVYPTQWFSDFTPFNPNMGIWKNWGLYFNQNIAPVWLGEFTTLQSTTDQWLKTLV 300  
 DB 241 ATSVYPTQWFSDFTPFNPNMGIWKNWGLYFNQNIAPVWLGEFTTLQSTTDQWLKTLV 300  
 QY 301 QYLRTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
 DB 301 QYLRTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 7  
 AAB48788 standard; protein; 521 AA.

XX AAB48788;  
 AC  
 XX 09-MAR-2001 (first entry)  
 DT  
 XX  
 DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y245G.  
 XX  
 KW El endoglucanase; glycosyl hydrolase; insoluble substrate;  
 KW cellulose hydrolysis; ethanol production; fermentation; mutant; muten.  
 XX  
 OS Acidothermus cellulolyticus.  
 OS Synthetic.  
 XX WO200070031-A1.  
 PN  
 XX 23-NOV-2000.  
 XX  
 PF 19-MAY-2000; 2000WO-US013971.  
 XX  
 PR 19-MAY-1999; 99US-0134925P.  
 XX  
 PA (MIDE ) MIDWEST RES INST.  
 XX

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;  
 PI Decker SR;  
 XX  
 DR WPI; 2001-061226/07.  
 XX  
 PT Preparation of glycosyl hydrolase with an increased catalytic activity on  
 PT insoluble substrate.  
 XX  
 XX Claim 5; Page 27-29; 30pp; English.

XX The invention relates to a method for making glycosyl hydrolase mutants  
 CC with increased catalytic activity with either insoluble or soluble  
 CC cellulose substrates relative to the wild-type enzyme. The method for  
 CC making a glycosyl hydrolase with increased insoluble substrate catalytic  
 CC activity comprises replacing an active site-associated amino glycosyl-  
 CC stabilising amino acid with an amino acid that does not strongly bind a  
 CC disaccharide product in the active site. Conversely, the method for  
 CC making a glycosyl hydrolase with increased soluble substrate catalytic  
 CC activity comprises replacing a hydrophobic substrate-binding amino acid  
 CC with a positively charged residue. The invention also discloses mutants  
 CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)  
 CC produced according to the method of the invention. The Y245G mutant  
 CC (AAB48788) has improved activity with insoluble substrates, and the W42R  
 CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with  
 CC soluble substrates. The invention also encompasses DNA encoding these  
 CC mutants. The glycosyl hydrolases of the invention are used as catalysts  
 CC for cellulose hydrolysis to produce sugars that can be fermented to  
 CC produce fuels such as ethanol. The present sequence represents the  
 CC Acidothermus cellulolyticus E1 endoglucanase Y245G mutant  
 XX

SQ Sequence 521 AA;

Query Match 99.1%; Score 1964; DB 4; Length 521;  
 Best Local Similarity 99.4%; Pred. No. 2.4e-178;  
 Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGYYHTSGREILDANNVPVRVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60  
 DB 1 AGGYYHTSGREILDANNVPVRVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60  
 QY 61 IRLPYSDILKPGTWPNSINFYQNMQLGLTSLOVMDKIYVAGQIGLRILDRHRPDC 120  
 DB 61 IRLPYSDILKPGTWPNSINFYQNMQLGLTSLOVMDKIYVAGQIGLRILDRHRPDC 120  
 QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 QY 181 AAERAGNAVLSVNPENLLIFVEGVQSYNGDSYWMGGLQAGQYPPVNLNPNRLVYSAHDY 240  
 DB 181 AAERAGNAVLSVNPENLLIFVEGVQSYNGDSYWMGGLQAGQYPPVNLNPNRLVYSAHDY 240  
 QY 241 ATSVYPTQWFSDFTPFNPNMGIWKNWGLYFNQNIAPVWLGEFTTLQSTTDQWLKTLV 300  
 DB 241 ATSVYPTQWFSDFTPFNPNMGIWKNWGLYFNQNIAPVWLGEFTTLQSTTDQWLKTLV 300  
 QY 301 QYLRTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
 DB 301 QYLRTAQYGADSFQWTFWSWNPDSGDTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 8  
 ADD24922  
 ID ADD24922 standard; protein; 535 AA.

XX  
 AC ADD24922;  
 XX  
 XX 15-JAN-2004 (first entry)  
 DT  
 XX  
 DE Xanthomonas campestris cellulase #3.  
 XX  
 KW Directed genetic engineering; galactonanase; reduced activity;  
 KW enhanced activity; xanthan gum production; suspension stability;



Db 282 ATSVYPTQWTFSDPTFPNNPGIWNKNGYLFNQNIAPVWLGFGTTLQSTDDQTLKTLV 341

QY 301 QYLRTAQYAGDSFQWTFWSPNPDSDGTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 358

Db 342 QYLRTAQYAGDSFQWTFWSPNPDSDGTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 399

RESULT 5

AA69508

ID AAY69508 standard; protein; 562 AA.

XX

AC AAY69508;

XX

DT 10-APR-2000 (first entry)

XX

DE Acidothermus cellulolyticus E1 endoglucanase.

XX

KW E1 endoglucanase; cellulose binding domain; CBD; cellulose modification;

XX beta-1,4-endoglucanase; endocellulase; thermostable.

XX

OS Acidothermus cellulolyticus.

XX

FH Key Location/Qualifiers

FT Peptide 1..41

FT /note= "Putative signal peptide"

FT Peptide 14..41

FT /note= "Putative signal peptide (alternative)"

FT Domain 42..404

FT /note= "Catalytic domain"

FT Region 405..460

FT /note= "Linker region"

FT Domain 461..562

FT /note= "Cellulose binding domain (CBD)"

XX

CA2226898-A1.

XX

PD 25-SEP-1999.

XX

PF 25-MAR-1998; 98CA-02226898.

XX

PR 25-MAR-1998; 98CA-02226898.

XX

PA (MIDE ) MIDWEST RES INST.

XX

PI Laymon RA, Adney WS, Thomas SR, Himmel ME;

XX

WPI; 2000-087663/08.

DR N-PSDB; AA255924.

XX

Isolated domains of Acidothermus cellulolyticus E1 endoglucanase useful for labelling or modifying a cellulose and for purifying or immobilizing a binding domain fusion protein to cellulose.

XX

Claim 1; Fig 2; 85pp; English.

XX

This sequence represents the Acidothermus cellulolyticus E1 endoglucanase, which is a beta-1,4-endoglucanase, or endocellulase. The cellulose binding domain (CBD) of E1 endoglucanase, and nucleotides which encode it are specifically claimed. The CBD is believed to be roughly wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge tip is inserted between the microfibrils of the cellulose fibre, disrupting the crystalline structure, and making the cellulose linkages more accessible to the catalytic domain of the E1 endoglucanase. The E1 endoglucanase CBD is useful in labelling or modifying the surface of cellulose or other polysaccharides. Such modified cellulose can then be used in textile, pulp, paper, chemical and pharmaceutical industries. CBDs can be used in affinity purification of CBD-fusion proteins, and can also be used to immobilise the CBD-fusion proteins to a cellulose support. CBD-fusion proteins can be used to modify the chemical or physical properties of a cellulose or polysaccharide matrix column and to modify (e.g., roughen or disrupt) a cellulose or polysaccharide fibre. The CBD of the E1 endoglucanase exhibits greater stability at pH 4-8 and has an optimum temperature for stability of 83 degrees Celsius which is

CC not found in CBDs from non-thermophilic organisms

XX

SQ Sequence 562 AA;

Query Match 99.4%; Score 1970; DB 3; Length 562;

Best Local Similarity 99.4%; Pred. No. 7.1e-179;

Matches 356; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRLMDQIKSLGYNT 60

Db 42 AGGGYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLSRDYSRLMDQIKSLGYNT 101

QY 61 IRLPYSDDLKPGTTPNSINFYQWQDQLGLTSLQWMDKIVAYAGQIGRIILDRHRPDC 120

Db 102 IRLPYSDDLKPGTTPNSINFYQWQDQLGLTSLQWMDKIVAYAGQIGRIILDRHRPDC 161

QY 121 SGQSALHTSSVSEATWISDLQALQRYKGNPVVGFDLHNEPHDPACWCGGSPSIDWRL 180

Db 162 SGQSALHTSSVSEATWISDLQALQRYKGNPVVGFDLHNEPHDPACWCGGSPSIDWRL 221

QY 181 AABRAGNAVLSVNFNLLIFVEGVQSYNGDSYMWGGLQAGQYFVVLNPNRLVYSAHDY 240

Db 222 AABRAGNAVLSVNFNLLIFVEGVQSYNGDSYMWGGLQAGQYFVVLNPNRLVYSAHDY 281

QY 241 ATSVYPTQWTFSDPTFPNNPGIWNKNGYLFNQNIAPVWLGFGTTLQSTDDQTLKTLV 300

Db 282 ATSVYPTQWTFSDPTFPNNPGIWNKNGYLFNQNIAPVWLGFGTTLQSTDDQTLKTLV 341

QY 301 QYLRTAQYAGDSFQWTFWSPNPDSDGTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 358

Db 342 QYLRTAQYAGDSFQWTFWSPNPDSDGTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 399

RESULT 6

AA648787

ID AAB48787 standard; protein; 521 AA.

XX

AC AAB48787;

XX

DT 09-MAR-2001 (first entry)

XX

DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y82R.

XX

KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;

XX cellulose hydrolysis; ethanol production; fermentation; mutant; muten.

XX

OS Acidothermus cellulolyticus.

OS Synthetic.

XX

PN WO200070031-A1.

XX

PD 23-NOV-2000.

XX

PF 19-MAY-2000; 2000WO-US013971.

XX

PR 19-MAY-1999; 99US-0134925P.

XX

PA (MIDE ) MIDWEST RES INST.

XX

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;

XX Decker SR;

XX

DR WPI; 2001-061226/07.

XX

PT Preparation of glycosyl hydrolase with an increased catalytic activity on insoluble substrate.

XX

PS Claim 17; Page 25-27; 30pp; English.

XX

The invention relates to a method for making glycosyl hydrolase mutants with increased catalytic activity with either insoluble or soluble cellulose substrates relative to the wild-type enzyme. The method for making a glycosyl hydrolase with increased insoluble substrate catalytic

XX 15-JUL-1994; 94US-00276213.  
 XX (MIDE ) MIDWEST RES INST.  
 XX Thomas SR, Laymon RA, Himmel ME;  
 XX WPI; 1996-105843/11.  
 XX N-PSDB; AAT12337.  
 XX New isolated DNA encoding endo:glucanase - obtd from Acidothermus  
 XX cellulolyticus, used for prodn of the enzyme for use in cellulose  
 XX hydrolysis.  
 XX Claim 1; Page 22; 34pp; English.

XX Acidothermus cellulolyticus E1 endoglucanase (AAR89927) is useful for  
 XX hydrolysing cellulosic biomass to sugars for simultaneous or subsequent  
 XX fermentation to ethanol. It shows optimal activity at 83 deg C. The amino  
 XX acid sequence of the mature enzyme was deduced from an E1 endoglucanase  
 XX gene (AAT12337) cloned from A. cellulolyticus. Putative signal peptides  
 XX were also identified (see also AAR89928-29). Cloning of this gene allows  
 XX large-scale, low-cost prodn. of recombinant E1 endoglucanase, using pref.  
 XX Saccharomyces, Zymomonas or E.coli hosts  
 XX Sequence 521 AA;

Query Match 99.6%; Score 1974; DB 2; Length 521;  
 Best Local Similarity 99.7%; Pred. No. 2.6e-179;  
 Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60  
 DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60  
 QY 61 IRLPYSDILKPGTNPNSINFYQMNQDLQGLTSQWMDKIVAYAGQIGRLILDRHRPDC 120  
 DB 61 IRLPYSDILKPGTNPNSINFYQMNQDLQGLTSQWMDKIVAYAGQIGRLILDRHRPDC 120  
 QY 121 SQGSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 121 SQGSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180  
 QY 181 AARAGNAVLNVNPNLLIFVEGVSYNGDSYVWGGNLQAGQYFVVLNVNPNRLVYSAHDY 240  
 DB 181 AARAGNAVLNVNPNLLIFVEGVSYNGDSYVWGGNLQAGQYFVVLNVNPNRLVYSAHDY 240  
 QY 241 ATSVYPTQWFSDFTPNNMPCGINWKNWGYLFNQNIAFPVWLGFGTTLQSTTDQTLWLT 300  
 DB 241 ATSVYPTQWFSDFTPNNMPCGINWKNWGYLFNQNIAFPVWLGFGTTLQSTTDQTLWLT 300  
 QY 301 QYLPTAQYCADSFQWTFWNSNPDSGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
 DB 301 QYLPTAQYCADSFQWTFWNSNPDSGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 4  
 ID AAU79549 standard; protein: 562 AA.  
 AC AAU79549;  
 XX 24-SEP-2002 (first entry)  
 DT A. cellulolyticus cellulase E1 beta-1,4-endoglucanase precursor.  
 DE Gene; ds; transgenic; plant; lignocellulose; cellulase; ligninase;  
 KW fermentable sugar; ethanol; fermentation; silage; feed; fuel;  
 KW industrial chemical; biodegradation; chloroaromatic;  
 KW environmental pollutant; E1 beta-1,4-endoglucanase; ei.  
 XX Acidothermus cellulolyticus.  
 OS  
 XX

FH Key Location/Qualifiers  
 FT Peptide l. 41  
 FT /label= Leader\_sequence  
 FT Protein 42. 562  
 FT /label= Mature\_cellulase

XX WO200234926-A2.  
 XX 02-MAY-2002.  
 XX 18-OCT-2001; 2001WO-US032538.  
 XX 20-OCT-2000; 2000US-0242408P.  
 XX (UNMS ) UNIV MICHIGAN STATE.

XX Sticklen WB, Dale BE, Maqbool S;  
 XX WPI; 2002-489947/52.  
 XX N-PSDB; ABK86729.  
 XX Producing transgenic plants which after harvest degrade lignin and  
 XX cellulose to fermentable sugars, by mating transgenic plant comprising  
 XX DNA encoding cellulase with transgenic plant comprising DNA encoding  
 XX ligninase.

XX Disclosure; Page 96-99; 126pp; English.

XX The invention discloses the production of a transgenic plant which  
 XX degrades lignocellulose when the plant is ground. It comprises the  
 XX production of the transgenic plant including cellulase and ligninase by  
 XX mating a transgenic plant, containing a DNA encoding a cellulase, and a  
 XX transgenic plant, containing a DNA encoding a ligninase, where both genes  
 XX are operably linked to a nucleotide sequence encoding a signal peptide  
 XX which targets the fusion protein to an organelle of the plant,  
 XX particularly chloroplasts. The method is useful for producing a  
 XX transgenic plant (e.g. maize) which degrades lignocellulose when the  
 XX plant is ground to produce a plant material. This material is useful for  
 XX converting lignocellulose, in a plant material, to fermentable sugars  
 XX which are then fermented to ethanol. The transgenic plants also provide a  
 XX plentiful and inexpensive source of fungal or bacterial cellulases and  
 XX ligninases which can be used in the production of ethanol. They can also  
 XX be used for pre-treating silage to increase the energy value of  
 XX lignocellulosic feeds for cows and other ruminant animals, pre-treating  
 XX lignocellulosic biomass for fermentative conversion to fuels and  
 XX industrial chemicals, and biodegradation of chloroaromatic environmental  
 XX pollutants. The protein sequence presented is the A. cellulolyticus  
 XX cellulase E1 beta-1,4-endoglucanase precursor

XX Sequence 562 AA;

Query Match 99.6%; Score 1974; DB 5; Length 562;  
 Best Local Similarity 99.7%; Pred. No. 2.9e-179;  
 Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 60  
 DB 42 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSDYRSMLDQIKSLGYNT 101  
 QY 61 IRLPYSDILKPGTNPNSINFYQMNQDLQGLTSQWMDKIVAYAGQIGRLILDRHRPDC 120  
 DB 102 IRLPYSDILKPGTNPNSINFYQMNQDLQGLTSQWMDKIVAYAGQIGRLILDRHRPDC 161  
 QY 121 SQGSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 162 SQGSALWYTSVSEATWISDLQALAQRYKGNPTVGFDLHNEPHDPACWCGDPSIDWRL 221  
 QY 181 AARAGNAVLNVNPNLLIFVEGVSYNGDSYVWGGNLQAGQYFVVLNVNPNRLVYSAHDY 240  
 DB 222 AARAGNAVLNVNPNLLIFVEGVSYNGDSYVWGGNLQAGQYFVVLNVNPNRLVYSAHDY 281  
 QY 241 ATSVYPTQWFSDFTPNNMPCGINWKNWGYLFNQNIAFPVWLGFGTTLQSTTDQTLWLT 300

CC (AA848788) has improved activity with insoluble substrates, and the W42R  
 CC (AA848786) and Y82R (AA848787) mutants have improved activity with  
 CC soluble substrates. The invention also encompasses DNA encoding these  
 CC mutants. The glycosyl hydrolases of the invention are used as catalysts  
 CC for cellulose hydrolysis to produce sugars that can be fermented to  
 CC produce fuels such as ethanol. The present sequence represents the  
 CC Acidothermus cellulolyticus E1 endoglucanase W42R mutant  
 XX  
 SQ Sequence 521 AA;

Query Match 100.0%; Score 1982; DB 4; Length 521;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-180;  
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNT 60  
 DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNT 60  
 QY 61 IRLPYSDDLKPGTTPMNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDC 120  
 DB 61 IRLPYSDDLKPGTTPMNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDC 120  
 QY 121 SGQSALWYTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 121 SGQSALWYTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGGLQAGQYPPVVLNPNRLVYSADHY 240  
 DB 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGGLQAGQYPPVVLNPNRLVYSADHY 240  
 QY 241 ATSVYPTQWFSDPTFPNNMGIWNKMGYLFNQNIAPIVWLGEFTTLQSTTDQWLKTLV 300  
 DB 241 ATSVYPTQWFSDPTFPNNMGIWNKMGYLFNQNIAPIVWLGEFTTLQSTTDQWLKTLV 300  
 QY 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
 DB 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 2  
 AAW39262  
 ID AAW39262 standard; protein; 358 AA.

XX AAW39262;  
 XX  
 XX 14-MAY-1998 (first entry)  
 XX  
 XX A. cellulolyticus E1-CAT translated region.  
 XX  
 XX E1 gene; endoglucanase; E1-CAT; catalytic domain; CD; enzyme; stability;  
 XX truncated; cellulose hydrolysis; biomass conversion.  
 XX  
 XX Acidothermus cellulolyticus.

XX US5712142-A.  
 XX  
 XX 27-JAN-1998.  
 XX  
 XX 22-FEB-1996; 96US-00604913.  
 XX  
 XX 26-SEP-1989; 89US-00412434.  
 XX  
 XX 27-JAN-1992; 92US-00826089.  
 XX  
 XX 21-SEP-1993; 93US-00125115.  
 XX  
 XX 15-JUL-1994; 94US-00276213.  
 XX  
 XX (MIDE ) MIDWEST RES INST.  
 XX  
 XX Chou Y, Himmel ME, Baker JO, Thomas SR, Adney WS;  
 XX  
 XX WPI; 1998-119985/11.  
 XX  
 XX N-PSDB; AAV09659.

XX DNA encoding truncated form of E1 endo-glucanase from Acidothermus  
 FT

PT cellulolyticus - and related vectors and transformed cells, expressing  
 PT only catalytic domain, is used for biomass conversion and has better heat  
 PT stability than complete enzyme.

XX Claim 2; Fig 5; 19pp; English.

XX This sequence represents a novel Acidothermus cellulolyticus endoglucanase  
 CC E1 protein (E1-CAT) which expresses only the catalytic domain (CD) of the  
 CC enzyme, without its peptide linker or cellulose binding domain (CBD) and  
 CC the truncated enzyme expressed by is used for hydrolysis of cellulose  
 CC (biomass conversion). Compared with full-length E1, the truncated enzyme  
 CC has better heat stability and higher temperature of maximum activity

XX Sequence 358 AA;

Query Match 99.6%; Score 1974; DB 2; Length 358;  
 Best Local Similarity 99.7%; Pred. No. 1.5e-179;  
 Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNT 60  
 DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLRSRDYRSMDOIKSLGYNT 60  
 QY 61 IRLPYSDDLKPGTTPMNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDC 120  
 DB 61 IRLPYSDDLKPGTTPMNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDC 120  
 QY 121 SGQSALWYTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 121 SGQSALWYTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGGLQAGQYPPVVLNPNRLVYSADHY 240  
 DB 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYMWGGLQAGQYPPVVLNPNRLVYSADHY 240  
 QY 241 ATSVYPTQWFSDPTFPNNMGIWNKMGYLFNQNIAPIVWLGEFTTLQSTTDQWLKTLV 300  
 DB 241 ATSVYPTQWFSDPTFPNNMGIWNKMGYLFNQNIAPIVWLGEFTTLQSTTDQWLKTLV 300  
 QY 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358  
 DB 301 QYLRPTAQYAGDSFQWTFWSNPDSDGTGGILKDDWQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 3  
 AAR89927  
 ID AAR89927 standard; protein; 521 AA.

XX AAR89927;  
 XX  
 XX 08-OCT-1996 (first entry)  
 XX  
 XX A. cellulolyticus E1 endoglucanase.

XX  
 XX E1 endoglucanase; cellulase; cellulose; saccharification; ethanol.  
 XX  
 XX Acidothermus cellulolyticus.

XX Key Location/Qualifiers  
 XX Domain 1..363  
 XX /label= Catalytic-domain  
 XX Domain 364..417  
 XX /label= Linker  
 XX /note= "proline/serine/threonine-rich linker domain  
 XX common to multi-domain microbial cellulases"  
 XX Domain 418..521  
 XX /label= Cellulose-binding\_domain

XX WO9602551-A1.  
 XX  
 XX 01-FEB-1996.  
 XX  
 XX 14-JUL-1995; 95WO-US008868.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:04 ; Search time 47.6667 Seconds  
(without alignments)  
2122.070 Million cell updates/sec

Title: US-09-997-504A-12

Perfect score: 1982  
Sequence: 1 AGGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1982	100.0	521	4	AAB48786
2	1974	99.6	358	2	AAW39262 A. cellul
3	1974	99.6	521	2	AAW89927 A. cellul
4	1974	99.6	562	5	AAU79549 A. cellul
5	1970	99.4	562	3	RAY69508 A. cellul
6	1965	99.1	521	4	AAB48787 A. cellul
7	1964	99.1	521	4	AAB48788 A. cellul
8	785.5	39.6	535	7	ADD24922 Xanthomon
9	759	38.3	628	2	AAW34999 Archaeabac
10	759	38.3	841	2	AAW34985 Archaeabac
11	758	38.2	553	5	ABG70759 T. mariti
12	751	37.9	430	3	ABR10344 P. horiko
13	751	37.9	458	3	ABR10345 P. horiko
14	694.5	35.0	514	4	AAW96264 Putative
15	617	31.1	616	2	RAY13494 Truncated
16	617	31.1	616	5	AAE16325 Active ce
17	617	31.1	1426	2	RAY13492 Truncated
18	617	31.1	1426	5	AAE16323 Active ce
19	509.5	25.7	425	4	ABR09060 Thermom
20	281.5	14.2	582	7	ADD24920 Xanthomon
21	267	13.5	508	5	ABR92443 Herbicida
22	205	10.3	38	2	AAW89930 A. cellul
23	205	10.3	389	5	AAW49427 Penicilli
24	204.5	10.3	488	5	ABR93510 Herbicida
25	189	9.5	551	5	ABR92442 Herbicida

26	181.5	9.2	762	6	ABP73022
27	179.5	9.1	375	6	ABP73019
28	176	8.9	517	6	ABP99336
29	154.5	7.8	722	6	ABU24069
30	151	7.6	420	6	ABP96833
31	149.5	7.5	329	2	AAW35002
32	146.5	7.4	518	4	ABR06928
33	146	7.4	666	2	AAW34992
34	145.5	7.3	406	2	AAW08199
35	142	7.2	431	3	AAW42172
36	142	7.2	431	3	AAW27501
37	142	7.2	431	5	ABR93972
38	142	7.2	442	3	AAW42171
39	142	7.2	443	3	AAW27500
40	141	7.1	335	7	ADD24924
41	141	7.1	400	4	AAW09785
42	138	7.0	375	3	AAW27502
43	138	7.0	375	3	AAW42173
44	138	7.0	562	4	AAW42094
45	138	7.0	562	6	ABR38613

ALIGNMENTS

RESULT 1

AAB48786  
ID AAB48786 standard; protein; 521 AA.

XX

XX AAB48786;

DT 09-MAR-2001 (first entry)

XX

DE Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.

XX

KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;

KW cellulose hydrolysis; ethanol production; fermentation; mutant; mutein.

XX

OS Acidothermus cellulolyticus.

OS Synthetic.

XX

PN WO200070031-A1.

XX

PD 23-NOV-2000.

XX

PF 19-MAY-2000; 2000WO-US013971.

XX

PR 19-MAY-1999; 99US-0134925P.

XX

XX (MIDE ) MIDWEST RES INST.

XX

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;

XX

PI Decker SR;

XX

XX WPI; 2001-061226/07.

DR

XX Preparation of glycosyl hydrolase with an increased catalytic activity on

XX insoluble substrate.

XX

PS Claim 16; Page 22-24; 30pp; English.

XX

CC The invention relates to a method for making glycosyl hydrolase mutants with increased catalytic activity with either insoluble or soluble cellulose substrates relative to the wild-type enzyme. The method for making a glycosyl hydrolase with increased insoluble substrate catalytic activity comprises replacing an active site-associated amino glycosyl-stabilising amino acid with an amino acid that does not strongly bind a disaccharide product in the active site. Conversely, the method for making a glycosyl hydrolase with increased soluble substrate catalytic activity comprises replacing a hydrophobic substrate-binding amino acid with a positively charged residue. The invention also discloses mutants of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788) produced according to the method of the invention. The Y245G mutant

QY 219 GAGQYVVLVNLVNLVYSAHDYATSVGPOTWFSDFTPFNMM-PGIWNKNWGVLEFNONIAP 277  
Db 440 GVADHPVVISADPKLVVSHVDGPDIYQWPKDFDINTLYEBCWYFNWYIYEQNIAP 499  
QY 278 VWLGEFGTTTQSTDTQWLKTLVQVLRPTAQYAGDSFQWTFWSPNPDSDGTGGILKDDWQ 337  
Db 500 MLIGEGWGLNENNRKNKLECLATEI-----AEKKLHHTFWAFNPNSADTGGMLMDWK 553  
QY 338 TVDTVTKDGYLAP 349  
Db 554 TVDEERYAIIVP 565  
RESULT 14  
Q95YQ0 PRELIMINARY; PRT; 534 AA.  
AC Q95YQ0; (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Cellulase Cel5-N.  
GN CELN  
OS Clostridium cellulolyticum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1521;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Belaich A., Gaudin C., Gal L.;  
RT "A gene cluster of cellulases of Clostridium cellulolyticum.";  
RL Submitted (CC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF316823; AAG4162.1; -.  
DR HSP; P54583; 1ECE.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR002105; Dockerin\_1.  
DR InterPro; IPR02048; EF-hand.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00404; Cellulase; 1.  
DR Pfam; PF00404; Dockerin\_1; 2.  
DR PROSITE; PS00018; EF\_HAND; 2.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_P5; 1.  
SQ SEQUENCE 534 AA; 59671 MW; 980785F0971ESA59 CRC64;  
Query Match 29.8%; Score 591.5; DB 2; Length 534;  
Best Local Similarity 32.8%; Pred. No. 1.1e-37;  
Matches 133; Conservative 60; Mismatches 144; Indels 69; Gaps 13;  
QY 6 W-HTSGREILDANNVPRVRIAGINWFGFTCNVYVHGLWSRDSRLDQIKSYNTIRLP 64  
Db 38 WLHCVGDKTYDMNGREVWLTGANWFGNCSNVFHGAW-YDVKNILTSVADRIGILLRVP 96  
QY 65 YSDIIL-----KPG-----TPNSINFYQNMQDLOG-----LTSQYMDKIVAYAGQI 107  
Db 97 ISTELLYSMWTKGNKVSSTASNNPPTVWNPDPFYDPATDGPKNSEIFDIIMKYKEL 156  
QY 108 GLRIILDRHPDCSQS-----ALWY-----TSVSEATWISDLQALQRYKGNPTVGF 158  
Db 157 GIKVMTDVHSPDANNSGHMYPLAYGLETGATTTDKWIDTLTLAGYKNDTILAI 216  
QY 159 LHNEPH-----DPACWCGDPSIDWRLAERAGNAVLSVNPNNLLIFVEGVQSY- 206  
Db 217 LKNEPHKRGYTNAAPTDMAKNDNTDENNKYAAERCSKEILAVNPKLLINIEGLEQYP 276  
QY 207 -----NGDSY-----WGGNLQAGQYPPVVLNVP-NRLVYSAHDYATSV 244  
Db 277 KTEKGYFTDTPVWAGSGDAAAPHWGGWGNLGRGVKDYPIDLGPLNSQIVYSPHYGPSV 336  
QY 245 GPOTWF-SDPTFPNNMFGIWNKNWGVLEFNONIAPVWLGEFGTTTQSTDTQWLKTLVQYL 303  
Db 337 YNQSWEFKDFTTTLDDYWDYTWAIYDQKLAFLIGEGWGGFMDGAKNKNWTLRLDYK 396

QY 304 RPTAQYAGDSFQWTFWSPNPDSDGTGGILKDDWQTVDTVKDGYLAP 349  
Db 397 I-----KXRNHTFWCLNPNNSGDTGGLIGNDSTWDEERYKGLKP 436  
RESULT 15  
Q7X2N2 PRELIMINARY; PRT; 616 AA.  
AC Q7X2N2; (TREMBlrel. 25, Created)  
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Endoglucanase.  
GN CEL5B.  
OS Thermomonospora fusca.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptosporangineae; Nocardiopsaceae; Thermobifida.  
OX NCBI\_TaxID=2021;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Posta K., Beki B., Kukolya J., Hornok L.;  
RT "Phylogenetic relationships of Tf cel5B, a new endoglucanase encoding  
RT gene from Thermobifida fusca.";  
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY298814; AAP56348.1; -.  
SQ SEQUENCE 616 AA; 67701 MW; 24FFC1EA1A3F5639 CRC64;  
Query Match 29.3%; Score 582; DB 2; Length 616;  
Best Local Similarity 34.3%; Pred. No. 7.4e-37;  
Matches 130; Conservative 62; Mismatches 149; Indels 38; Gaps 13;  
QY 1 AGGYW-HTSGREILDANNVPRVRIAGINWFGFTCNVYVHGLWSRDSRLDQIKSLGYN 59  
Db 33 SGTADWLHTDGNRIVDSAGNEVWLTGANWFGFTSERMFHGLWAANTIEDITSAMAERGIN 92  
QY 60 TIRLPYSDDIL---KPGTM-PNSINFYQNMQDLOGTSLQVMDKIVAYAGQIGLRIILDR 115  
Db 93 MVRVPISTQLLEWKNQAGSPGVNEY-VNPELAGMNTLEVFYWLQICEYGLKWLMDV 151  
QY 116 H--RPDCSGQ-SALWYTSVSEATWISDLQALQRYKGNPTVYGFDLHNEPHDP----- 166  
Db 152 HSAEADNSGHVYVWYKGDITTEDFYTAWEWVTERYKNDTIVAADIKNEPHGKANETPR 211  
QY 167 ACWCGDPSID-WRLAERAGNAVLSVNPNNLLIFVEGVQSYNGD-----SYW 212  
Db 212 AKWD-GSTIDNFKVCETAGKRLAINPNMLLCEGIEIYPKDQGWSDTGRDYISTW 270  
QY 213 WGGNLQAGQYPPVVLNV-PNRLVYSAHDYATSVGPOTWFSDFTPFNMM-PGIWNKNWGV 270  
Db 271 WGGNLRGVADHPVDLGAHQDQVYSPHYGPSVFEQWPFGEWNRQTLTDEWVRPNWLYI 330  
QY 271 PNQIAPVWLGEFGTTTQSTDTQWLKTLVQVLRPTAQYAGDSFQWTFWSPNPDSDGTGG 330  
Db 331 HEDDIAFLIGEGWGGFLDGDGNEKMTALRSIID-----EKMHHTFWALNPNNSGDTGG 384  
QY 331 ILKDDWQTVTKDGYLAP 349  
Db 395 LLNYDWTWDEAKYAFKP 403  
Search completed: August 2, 2004, 16:36:56  
Job time : 45.3333 secs

829 RGVRYKYPINLGKYNKVVSPHSDYGVSPVQOPWFY-PGFTKESLLODCWRPNWAYIMEEN 887  
Db  
275 IAPVWLGEFTGTLQSTTDQWLKTLVQYLRPTAQYGADSFQWTFWSNPDSDGTGILK 334  
Qy  
888 IAPLLIGEWGHLGDGADNEXKWKYLRDYII-----ENHIIHTFWCFNANSBGDTGLGVY 941  
Db  
335 DWQVTVTKDGYLAP 349  
Qy  
942 DFTTWDEKYSFLKP 956  
Db  
RESULT 12  
ID Q9X3P6 PRELIMINARY; PRT; 1426 AA.  
AC Q9X3P6  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CelB.  
GN CelB.  
OS Caldicellulosiruptor sp. Tok7B.1.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Syntrophomonadaceae;  
OC Caldicellulosiruptor.  
OX NCBI\_TaxID=80339;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TOK7B.1;  
RX MEDLINE=20171169; PubMed=10706665;  
RA Gibbs M.D., Reeves R.A., Farrington G.K., Anderson P., Williams D.P.,  
RA Bergquist P.L.;  
RT "Multidomain and multifunctional glycosyl hydrolases from the extreme  
thermophile Caldicellulosiruptor isolate Tok7B.1.";  
RL Curr. Microbiol. 40:333-340(2000).  
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL  
HYDROLASES).  
CC EMBL; AF078737; AAD30364.1; -.  
DR HSSP; Q06851; 1N8C.  
DR GO; GO:0003779; F:actin binding; IEA.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR InterPro; IPR001589; Actbind actnin.  
DR InterPro; IPR001956; CBD 3.  
DR InterPro; IPR008965; Cellul bind.  
DR InterPro; IPR010000; Glyco\_hydro.10.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00942; CEM\_3; 3.  
DR Pfam; PF00150; cellulase; 1.  
DR Pfam; PF00331; Glyco\_hydro.10; 1.  
DR PRINTS; PR00134; GLYDRLASE10.  
DR ProDom; PD001947; CBD 3; 3.  
DR SMART; SMO0633; Glyco.10; 1.  
DR PROSITE; PS00019; ACTININ 1; 1.  
DR PROSITE; PS00591; GLYCOSYL\_HYDROL\_F10; 1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
KW Glycosidase; Hydrolase.  
SQ SEQUENCE 1426 AA; 157544 MW; 29B3FDB85D09A863 CRC64;  
Query Match 31.3%; Score 621; DB 2; Length 1426;  
Best Local Similarity 35.5%; Pred. No. 2.1e-39;  
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;  
Qy 10 GREILDANNVVRVIRAGINWFGFTCTN-VVHGLWSRDYRMLDQIKSLGYNILRPVSDDI 69  
Db 1022 GNKIVDKGKFWLTVGNWFGFTCTN-VVHGLWSRDYRMLDQIKSLGYNILRPVSDDI 1081  
Qy 70 L--KPGTMEP-SINFYOMNODLOGLTSLQWMDKIVAYAGQIGLRIILDRH--R 123  
Db 1082 ILNWSKGLIYKPNINYN-VNPELSEGLTSLQWMDKIVAYAGQIGLRIILDRH--R 1140  
Qy 124 -SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHD-----ACWCGDPS 175  
Db 1141 IYVWYTDITTPEDYKACEWITERYKNDTIVAFDLKNEPFGKWPQDSVFAKWDNSTD 1200

176 IDWRLAERAGNAVLSVNPILLIFVEGVQSYNGP-----SYWNGNLOQAGQ 222  
Qy  
1201 NNWYAAETCAKRIILAKNPMLIVIEGLEAYPKDDVTWTSKSSDYISTWGGNLRGUKK 1260  
Db  
223 YPVVL-NVNPRLVTSADHYATSVGPQTW-----FSDPTFPNNMFGIMKNKMGYLFNQNIAP 277  
Qy  
1261 YPINLGQYQNVKVPSPHDYGLVYQOPWFYPGFTKDTLYND---CWRDNWYIMDNGIAP 1317  
Db  
278 VWLGEFGTTLQSTTDQWLKTLVQYLRPTAQYGADSFQWTFWSNPDSDGTGILK 337  
Qy  
1318 LLIGEWGGLDGDGNEKWMYLRDYII-----ENHIIHTFWCFNANSBGDTGLGVYD 1371  
Db  
338 TVDTVTKDGYLAP 349  
Qy  
1372 TWDEKYNFLAP 1383  
Db  
RESULT 13  
ID Q9L3J2 PRELIMINARY; PRT; 660 AA.  
AC Q9L3J2  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE 1.4-beta-cellobiohydrolase.  
GN CELO.  
OS Clostridium thermocellum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1515;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=F7;  
RA Zverlov V.V., Schwarz W.H.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBDJ databases.  
DR EMBL; AJ275975; CAB76938.1; -.  
DR HSSP; P54583; 1ECE.  
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.  
DR GO; GO:0004678; F:hydrolase activity; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR008965; Cellul bind.  
DR InterPro; IPR002105; Dockerin.1.  
DR InterPro; IPR001547; Glyco\_hydro.5.  
DR InterPro; IPR000437; Prok\_lipoprot\_s.  
DR Pfam; PF00150; cellulase; 1.  
DR Pfam; PF00404; Dockerin.1; 2.  
DR PROSITE; PS00448; CLOS\_CELLULOSOME\_RPT; 1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Hydrolase.  
SQ SEQUENCE 660 AA; 75247 MW; 04A042002C288CFC CRC64;  
Query Match 31.2%; Score 620.5; DB 2; Length 660;  
Best Local Similarity 35.8%; Pred. No. 8.2e-40;  
Matches 133; Conservative 66; Mismatches 138; Indels 35; Gaps 10;  
Qy 6 W-HTSGREILDANNVVRVIRAGINWFGFTCTN-VVHGLWSRDYRMLDQIKSLGYNILRL 63  
Db 201 WLVHVEGLIKDAQQNTVYLTGINWFGFTCTN-VVHGLWSRDYRMLDQIKSLGYNILRL 260  
Qy 64 PYSDIL--KPGTWPNS--INFYOMNODLOGLTSLQWMDKIVAYAGQIGLRIILDRH-R 117  
Db 261 PLASIIILQWNGRVERVTSFVNTYE-NPRLDGLSLEILDYTNHMKNGKAMKIDWESS 319  
Qy 118 PDCSQSALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPH-----DPACWGC 171  
Db 320 TKDSYQENLWYKNDITMEEFIEAWKWIVERYKDDTIVAFDLKNEPFGKYSGPNIKWDD 379  
Qy 172 GPSIDWRLAERAGNAVLSVNPILLIFVEGVQSYNGDSY-----WVGGLQ 218  
Db 380 SNDPNNWKAELIIEELIATNPILLIVVEGVEATPMEGDYTCGFTTYCNWGGNLR 439

QY 236 SAHDYATSGPOTWFSDFPTFNNMPGINKWGLYFNQNIAPVWLGFGTTL--QSTTDQ 293  
 Db 258 MPHVGPDVYVQFYFNSPDPFNNAALWDKHFHAKAGYA-MAIGFSGYKGEGRDI 316  
 QY 294 TWLXTLVQYLRPTAQY-ADSFQWTFWMSNPDSDTGGILKDDWQTV--DTVK 343  
 Db 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSVDITGMYGNDWTFPRDDKVK 363

RESULT 11  
 ID 024820 PRELIMINARY; PRT; 1000 AA.  
 AC 024820;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Beta-glucanase.  
 OS thermophilic anaerobe NA10.  
 OC Bacteria.  
 OX NCBI\_TaxID=67756;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NA10;  
 RA Miyake K., Machida Y., Hattori K., Iijima S.;  
 RT "Characterization of a multi-domain cellulase from an extremely  
 thermophilic anaerobe strain NA10."  
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY F (FAMILY 10 OF GLYCOSYL  
 HYDROLASES).  
 DR EMBL; AB080029; BAA22939.1; -.  
 DR HSP; Q08851; INBC.  
 DR GO; GO:0003779; F:actin binding; IEA.  
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001589; Actbind\_actnin.  
 DR InterPro; IPR001956; CBD\_3.  
 DR InterPro; IPR001000; Glyco\_hydro\_10.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR Pfam; PF00942; CBM\_3; 1.  
 DR Pfam; PF00150; cellulase; 1.  
 DR Pfam; PF00331; Glyco\_hydro\_10; 1.  
 DR PRINTS; PR00134; GLHYDRLASE10.  
 DR ProDom; PD001947; CBD\_3; 1.  
 DR SMART; SM00633; Glyco\_10; 1.  
 DR PROSITE; PS00019; ACTININ\_1; 1.  
 DR PROSITE; PS00591; GLYCOSYL\_HYDROL\_F10; 1.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 KW Glycosidase; Hydrolase.  
 SQ SEQUENCE 1000 AA; 113265 MW; B9F659A56A752C6B CRC64;

Query Match 31.6%; Score 627.5; DB 2; Length 1000;  
 Best Local Similarity 35.5%; Pred. No. 4.1e-40;  
 Matches 133; Conservative 69; Mismatches 134; Indels 39; Gaps 12;

QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDSRMLDQIKSLGYNIRLP 64  
 Db 590 WLYVSGNKIVDKGRPVWLGVNWFNTGTNVPDGVWSNULSTLAIEANRGNLRVP 649  
 QY 65 YSDILG---RPGTTPN-SINFYQMNODQLGTLQVMDKIVAYAGQIGLRIILDRH--RP 118  
 Db 650 ISABELLWNSQGIYPKENINY-VNPELEGKNSLEVDIVVQTCKEVGLKIMLDHSIKT 708  
 QY 119 DCSQ-SALWYTSVSEATWISDLQALAQYKGNPTVGVFDLNEHPD-----ACWG 170  
 Db 709 DAMGHIPVWDEFKTFEDFKACEWTNRYKNDTTIIAFDLKNEPHKQWQDTTFKWD 768  
 QY 171 CGDPSIDRWLAARAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWGGNLL 217  
 Db 769 NSTDINNWKYAAETCAKRIINPNLLIVIEGIEAYPKDDVTWTSKSSDYSTWGGNLL 828  
 QY 218 QGACQYFVWL-NVFNRLVYSAHDYATSGPOTWFSDFPTFNN--MPCGIWKNWGLYFNQNI 274

QY 288 QSTTDQWTLVQYLRPTAQYAGDSFQWTFWMSNPDSDTGGILKDDWQTVK 343  
 Db 326 EG-LDKWQDAFYKWLIIKKIY-----NFFYCLNPESGDTGGIFLDDWKTYNWEK 375

RESULT 10  
 ID 087AH4 PRELIMINARY; PRT; 614 AA.  
 AC 087AH4;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Endo-1,4-beta-glucanase.  
 GN ENXCA OR PD181.  
 OS Xylella fastidiosa (strain Temecula / ATCC 700964).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 OX NCBI\_TaxID=183190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22421331; PubMed=12533478;  
 RA Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,  
 Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,  
 Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,  
 Goldnan M.H.S., Goldman G.H., Lemos M.V.F., El-Dorri H., Tsai S.M.,  
 Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,  
 Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuranae E.E.,  
 Marino C.L., Gigliotti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,  
 Baia G.S., Bianco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,  
 da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,  
 Leoni S.G., Oliveira A.R., Rosa V.B. Jr., Sassaki F.T., Sena J.A.D.,  
 de Souza A.L., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,  
 Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,  
 Kitajima J.P.;  
 RA "Comparative analyses of the complete genome sequences of Pierce's  
 disease and citrus variegated chlorosis strains of Xylella  
 fastidiosa";  
 RT J. Bacteriol. 185:1018-1026(2003).  
 RL EMBL; AB012560; AAQ29683.1; -.  
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.  
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
 DR InterPro; IPR001919; Bac\_Cellose-bind.  
 DR InterPro; IPR008965; Cellul\_bind.  
 DR InterPro; IPR002952; Eggshell.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR Pfam; PF00553; CBM\_2; 1.  
 DR Pfam; PF00150; cellulase; 1.  
 DR PRINTS; PR01228; EGGHELL.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 614 AA; 62041 MW; 29EC67B42F4BC3F7 CRC64;

Query Match 32.7%; Score 649.5; DB 16; Length 614;  
 Best Local Similarity 39.9%; Pred. No. 4.1e-42;  
 Matches 141; Conservative 58; Mismatches 125; Indels 29; Gaps 12;

QY 5 YHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDSRMLDQIKSLGYNIRLP 64  
 Db 26 YSISHGKVDWDDKN-QIQLRGVNWFGFETGHDVYVGLWLNWKNKEFITQLQGGFNIRLP 84  
 QY 65 YSDILKPGTTPN-SINFYQMNODQLGTLQVMDKIVAYAGQIGLRIILDRHPCSGOS 124  
 Db 85 FCPANLNSYSPSID-YGRNPDQLGSLQILDKVKLSDRMYVLDRHPCSAIS 143  
 QY 125 ALWYTSVSEATWISDLQALAQYKGNPTVGVFDLNEHPDPAFCGCGDPSIDWLAER 184  
 Db 144 ELWYTSYSEKQWIDLRFAHYRYANVHVGIVGLEVKNEPHGRITWGTGPKTDMNTAVEH 203  
 QY 185 AGNAVLNVNPNLLIFVEGVQSYNGDS-----YWWGGNLOGAGYPPVLANP-NRLVY 235  
 Db 204 AAMAILAARPKWLLIGVEGI-----GENPICSSTIGHFWGENLEPMDCTP--LKVPAHLL 257

```

Db 377 TIDTKKLALVQP 388
RESULT 8
Q9PF60 PRELIMINARY; PRT; 592 AA.
AC Q9PF60;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Endo-1,4-beta-glucanase.
GN XP0818.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=945C.
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia J.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Fromme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.B., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miranda E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.B.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R.C., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Teai S.M., Tsuchiko M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE003921; AAF83628.1; -.
DR PIR; B2759; E82759.
DR HSSP; P54583; 1ECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 592 AA; 59967 MW; 9846DA4EA3B5C93E CRC64;

Query Match 34.6%; Score 687.5; DB 16; Length 592;
Best Local Similarity 42.6%; Pred. No. 4.3e-45;
Matches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMQLDKSLGYNTRLRP 64
Db 26 YSIHGKVIDDKGN-QIQLKGSWFGFETTNVHGLWTRNWKKEFFIQISMGNAVRLP 84
QY 65 YSDDLKFGTNPENSYNQNDLQGLTSIQMDKIVAYAGQIGLRILDRHPDCSQS 124

```



```
Db 262 YGPDVFVQSVFNDSPNNMPAIDRHFQCPAGSH--ALLLGFQCKYEGGDARDKWKQD 319
QY 298 TLVQVLRPTAQYAGDSFQWTFWSNPDSDGTGILKDDWQTV 339
Db 320 ALVKYLR---SKGIN--EGFYWSNPNPNSGDTGGLRDDWTSV 356

RESULT 6
OS8925 PRELIMINARY; PRT; 458 AA.
AC OS8925;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 458AA long hypothetical endo-1,4-beta-glucanase.
GN PH1171.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3.
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamanoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb P.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.
PIR: E97012; E97012.
DR GO: 0005509; P:calcium ion binding; IEA.
DR GO: 0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO: 0005975; P:carbohydrate metabolism; IEA.
DR GO: 0005975; P:carbohydrate metabolism; IEA.
DR GO: 0000272; P:polysaccharide catabolism; IEA.
DR InterPro: IPR002105; Dockerin 1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001547; Glyco_hydro_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00404; Dockerin 1; 2.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 458 AA; 51930 MW; CAE48AD43A8EE654 CRC64;

Query Match 38.0%; Score 755; DB 17; Length 458;
Best Local Similarity 45.0%; Pred. No. 1.7e-50;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSREILDANNVPVRIAGINWGFETCNVYVHGLWSRDYRSLMDQTKSLGYNTRILPYSD 67
Db 54 TSGE-----TPIHLFGVNWFGFETPNVHGLWKRNMEDMLLQLKSLGFNAIRLPFT 107
QY 68 DILKPGTNPNSINFYQMNQDQLGLTSQVMDKIVAYAGIQLRIILDRHRPDCSQSALW 127
Db 108 ESKVPGTQPIGID-YSKNPDRLGLSLQIMEKIIKKAGDLGIFVLVDYHRICTHIEPLW 166
QY 128 YTSVSEATWISDLQALQRYKGNPTVGFDLNHPHD-----PACWGGDPS 175
Db 167 YTEDFSEEDFINTWIEVAKRFKGVNVIAGDLKNEPHSVTSPPAAVTDGTGATWNGNEPA 226
QY 176 IDWRLAERAGNAVLNPNLLIFVEGVQSYN-----GDSYVWNGNLOQAGQYPVY 226
Db 227 TDNLAERIGKAILKAVPHLIFVEGTQPTNPDKTSSYKMGYNAMWGNLMAVKDYPV- 285
QY 227 LNVPL-NRLVYSADHYATSVCQPTWFSDDPT-FPNMFGINKWNGYLFNQNIAPVWLGEG 284
Db 286 -NLPRNKLVTSPHYGPDVYQYFGAKGFPDNLFDIWHYHFGYKLELGSYVVIQKNIAP 344
QY 285 TTL---QSTTDQWLKTLVQYLRPTAQYAGDSFQWTFWSNPDSDGTGILKDDWQTV 339
Db 345 GKYGCGDPRDVIWQNKLVDMW---IENKFCDF---YWSNPNPNSGDTGILQDDWTTI 397
```

## RESULT 7

```
Q97KK6 PRELIMINARY; PRT; 482 AA.
ID Q97KK6
AC Q97KK6;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Possible non-processive endoglucanase family 5, secreted, Cella
DE homolog secreted, dockerin domain.
GN CAC0912.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
PIR: E97012; E97012.
DR GO: 0005509; P:calcium ion binding; IEA.
DR GO: 0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO: 0005975; P:carbohydrate metabolism; IEA.
DR GO: 0005975; P:carbohydrate metabolism; IEA.
DR GO: 0000272; P:polysaccharide catabolism; IEA.
DR InterPro: IPR002105; Dockerin 1.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001547; Glyco_hydro_5.
DR Pfam: PF00150; cellulase; 1.
DR Pfam: PF00404; Dockerin 1; 2.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Complete proteome.
SQ SEQUENCE 482 AA; 54109 MW; 13E2E9197D59A5F CRC64;

Query Match 34.7%; Score 690; DB 16; Length 482;
Best Local Similarity 37.9%; Pred. No. 2.1e-45;
Matches 141; Conservative 71; Mismatches 126; Indels 34; Gaps 9;

QY 5 YWTSGREILDANNVPVRIAGINWGFETCNVYVHGLWSRDYRSLMDQTKSLGYNTRILP 64
Db 24 YLHSDGSKLLDDYGNQVRMTGIAWFGLETPTNYCFHGLWANRLDNLINIVADNGFNTLRVP 83
QY 65 YSDDIL---KPGT--MENSINFYQMNQDQLGLTSQVMDKIVAYAGIQLRIILDRHRPD 119
Db 84 LSVELVNWQRGVPTPDSINDY-ISPELKGQNSLQILDVIAYSKVGKVMMDMERIE 142
QY 120 CSGSALWYTSVSEATWISDLQALQRYKGNPTVGFDLNHPHDPA-----WGCG 172
Db 143 SGGQTATWYTSKYTTDDYEKCWQYLDARYKNDTDTVIAADIFNEPHGKAYRAETSAKWN 202
QY 173 DPSIDWRLAERAGNAVLNPNLLIFVEGVQSYN-----NGDSY---WNGNLOQ 219
Db 203 TDEDNRYEASKVKKILIDINPKMLIVVEGVETTPKEGTAGSTNPDDYVGGWGGNLRG 262
QY 220 AGQYPPVL-NVNPRLVYSADHYATSVCQPTWTF--SDTFFNNMFGINKWNGYLFNQNIAP 277
Db 263 VKDYPVDLAPYKQNVVYSPHDYGPVSDQPTWFDGDFTEQSLNDIWRPSWFIQKNIAP 322
QY 278 VWLGEFGTTLQSTTDQWLKTLVQYLRPTAQYAGDSFQWTFWSNPDSDGTGILKDDWQ 337
Db 323 LLIGEWGNGMDGKQNEQMTDMAKLI-----SDKNMHTFWCLNANSNGDTGGILEYDFK 376
QY 338 TVDVTWKGYLAP 349
```

```

Db      165 SELWYTYSEYPERSWISDRMLAKRYASDPTVIGVDLHNEPHGAATWGTGAATTDWRAAE 224
Qy      184 RAGNAVLAVNPNLLIFVEGV-QSYNGDSYWMGNLQAGQYPPVVLNPNRLVYSAHDYAT 242
Db      225 RGNNAVLAEKPKLVIVGEGIDHQADGTGTWGGALDSAAATASVRLTVANRVVYSPHYDPS 284
Qy      243 SVGQOTWFSPTFPNNMGPKNWKGVLFNQNTAPWVLGFGCTTLOSTTQWTKTLVQY 302
Db      285 TIYGQPFASNYPTNLPGIWDHAGYLAKKDITAPVLVGEFGTKLETASDKQMLNTLVGY 344
Qy      303 LRPTAQGADSFQWTFWMPDSDGTGGILKDDWQTVDTVKDGYLAPI 350
Db      345 LSST--GISS--SFAWFPDSDGTGGIVKSDWVPEQAQKLDALAPI 386

RESULT 4
Q9EVR6 PRELIMINARY; PRT; 518 AA.
AC Q9EVR6
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Extracellular endoglucanase (ENGXCA protein) precursor.
GN ENGXCA.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RC STRAIN=NRL-B1459;
RX Schroter K., Puchler A., Becker A.;
RT "engXCA major extracellular endoglucanase.";
RL Submitted (DSC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ304415; CAC18529.1; -.
DR HSSP; P54583; 1ECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR SIGNAL.
FT CHAIN 1 25 POTENTIAL.
FT CHAIN 26 518 EXTRACELLULAR ENDOGLUCANASE (ENGXCA
FT PROTEIN).
SQ SEQUENCE 518 AA; 55588 MW; 9A0FF678E3BF712C CRC64;

Query Match 40.3%; Score 800.5; DB 2; Length 518;
Best Local Similarity 47.2%; Pred. No. 5.6e-54;
Matches 162; Conservative 58; Mismatches 104; Indels 19; Gaps 9;

Qy 5 YWHTSGEILDANNVPVRIAGINWFGFETCYVYVHGLWSRDYRSLDQIKSLGVNTRLP 64
Db 26 YSINNSQIVDDSKVQLKGVNVFGETGNVHVGHWGLWARKWMDVMQVGLGFNAVLRF 85
Qy 65 YSDILKPGTMPSINFYQMNQDLQGLTSLQVMDKIYAYAGQIGLRIILDRHPCDSCGOS 124
Db 86 FCPATLRSDTMFASID-YSRNADLQGLTSLQILDKVIAEFNARGMYVLLDHTPDCAGIS 144
Qy 125 ALWYTSVSEATWISDLOALAQRYKGNPTVYVGLDHLNEPHDPCWCGDPSIDWRLAAER 184
Db 145 ELWYTSVTEAQWLDRFVANRYKNTVYVGLDLKNEPHGAATWGTGNATDWNKRAER 204
Qy 185 AGNAVLSPNPLLIIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVVLNVP-NRLVYSAH 238
Db 205 GSAVLAVAPKWLIAVEGIDTNPVCSTNG-GIPWGNLQPLACTP--LNIPANRLLIAPH 261
Qy 239 DYATVSGPOTWFSPTFPNNMGPKNWKGVLFNQNTAPWVLGFGCTTLOSTTQWTKTLVQY 296
Db 262 VYGPDPVYVQSYFNSFNPNMGPKNWKGVLFNQNTAPWVLGFGCTTLOSTTQWTKTLVQY 319
Qy 297 KTLVQYLRPTAQYAGDSFQWTFWMPDSDGTGGILKDDWQTV 339

```

```

Db      320 DALVKYLK---SKGIN--QGFYWSWNPNSGDTGGILRDDWTSV 357
Qy      320 DALVKYLK---SKGIN--QGFYWSWNPNSGDTGGILRDDWTSV 357

RESULT 5
Q8PPS3 PRELIMINARY; PRT; 474 AA.
AC Q8PPS3
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Cellulase.
GN ENGXCA OR XAC0612.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A., L.P.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.P., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL; AB011689; AAM35501.1; -.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001519; Bact_cellulose_bind.
DR InterPro; IPR008965; Cellul_bind.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Complete proteome.
SQ SEQUENCE 474 AA; 51294 MW; 379B84D63CASA31B CRC64;

Query Match 40.2%; Score 799.5; DB 16; Length 474;
Best Local Similarity 46.5%; Pred. No. 5.9e-54;
Matches 159; Conservative 60; Mismatches 104; Indels 19; Gaps 9;

Qy 6 WHTSGEILDANNVPVRIAGINWFGFETCYVYVHGLWSRDYRSLDQIKSLGVNTRLP 65
Db 26 YSINNSQIVDDSKVQLKGVNVFGETGNVHVGHWGLWARKWMDVMQVGLGFNAVLRF 85
Qy 66 YSDILKPGTMPSINFYQMNQDLQGLTSLQVMDKIYAYAGQIGLRIILDRHPCDSCGOSA 125
Db 86 FCPATLRSDTMFASID-YSRNADLQGLTSLQILDKVIAEFNARGMYVLLDHTPDCAISE 144
Qy 125 ALWYTSVSEATWISDLOALAQRYKGNPTVYVGLDHLNEPHDPCWCGDPSIDWRLAAERA 185
Db 145 ELWYTSVTEAQWLDRFVANRYKNTVYVGLDLKNEPHGAATWGTGNATDWNKRAERG 204
Qy 186 AGNAVLSPNPLLIIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVVLNVP-NRLVYSAH 239
Db 205 GSAVLAVAPKWLIAVEGIDTNPVCSTNG-GIPWGNLQPLACTP--LNIPANRLLIAPH 261
Qy 240 YATVSGPOTWFSPTFPNNMGPKNWKGVLFNQNTAPWVLGFGCTTLOSTTQWTKLW 297

```

```
QY 124 SALWYTSVSEATWISDICALAQRKGNPTVGVFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 156 SELWHTSQYPSERWISDWMKLAERYKNNPTVIGADLHNEPHGQAWGTDVSTWRLAAQ 215
QY 184 RAGNAVLSVNPVNLIFVEGVQ---SYNGDSYWMGNGLAGAGQYPVVLANVNPRLVYSADHY 240
Db 216 RAGNAVLSVNPVNLIFVEGVQDHNVKGNNSQYWMGNGLAGVANYPVVLDVNPVNVVSPHDY 275
QY 241 ARSVGQTFWSDPTFPNMPGIIWKNWGYLFNONTAPVWLGEFG--FTLQSTTDOTWLT 298
Db 276 GFGVSSQPFNDSTFPNLPALWDQTWGYISKNTAPVILVGEFGGRNVLDSSPEGKQWNA 335
QY 299 LVQYLRLPTACYGADSFQWTFWSWNPDSGDTGILKDDQVTDVTKDGYLAPIKSIFFPV 358
Db 336 LVDYI-----GANNLYFIWNLNSGDTGILLDDWTTWNRPKQDML-----SRIMKPV 385

RESULT 2
Q9K5C7 PRELIMINARY; PRT; 745 AA.
ID Q9K5C7
AC Q9K5C7
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DE 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Cellulase precursor (Fragment).
GN CELA.
OS Clavibacter michiganensis.
OG Plasmid pCMI.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=28447;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Meleuzus D.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-614 FROM N.A.
RC STRAIN=NCPPB 382;
RX MEDLINE=20331586; PubMed=10875331;
RA Jahn H.; Dreier J.; Meleuzus D.; Bahro R.; Eichenlaub R.J.;
RT "The endo-beta-1,4-galactanase CelA of Clavibacter michiganensis subsp.
RT michiganensis is a pathogenic determinant required for induction of
RT bacterial wilt of tomato.";
RL Mol. Plant Microbe Interact. 13:703-714(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NCPPB 382;
RA Jahn H.;
RL Submitted (OCT-1991) to the EMBL/GenBank/DBJ databases.
DR ENBL; X62582; CAA44467.2; -.
DR HSSP; P54583; LECE.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KW Plasmid; Signal.
FT SIGNAL 66 POTENTIAL.
FT CHAIN 67 >745 CELLULOASE.
FT NON_TER 745
SQ SEQUENCE 745 AA; 77886 MW; 5B45015EB8D4F27C CRC64;

Query Match 45.4%; Score 901.5; DB 2; Length 745;
Best Local Similarity 50.0%; Pred. No. 1.3e-61;
Matches 174; Conservative 51; Mismatches 112; Indels 11; Gaps 4;
```

```
QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMDOI KSLGYNTIRL 63
Db 70 GWLHTAGGKIIVTASGAPYIRGIAMFGMETSSCAPHGLDTITLGGMQHILKQMGFTTVRL 129
QY 64 PYSDILKPGTNPNSINFYQNMQDLOGLTSLQWMDKIVAYAGQIGLRIILDRHRPDCSGQ 123
Db 130 PFSNQCL-----AASGVTVGSADPSLAGLTPLOVMDHVVASAKSAGLDVILDQHRPDSGGQ 185
QY 124 SALWYTSVSEATWISDICALAQRKGNPTVGVFDLHNEPHDPACWCGDPSIDWRLAAE 183
Db 186 SELWHTSQYPSERWISDWMKLAERYKNNPTVIGADLHNEPHGQAWGTDVSTWRLAAE 245
QY 184 RAGNAVLSVNPVNLIFVEGV-QSYNGDSYWMGNGLAGAGQYPVVLANVNPRLVYSADHYAT 242
Db 246 RGNNAVLAENFNLVLVEGIDHEADSGTWWGAGLGVGNFVRLSVANRVVSPHDYPS 305
QY 243 SVGPOTWFSPTFPNMPGIIWKNWGYLFNONTAPVWLGEFGTTLQSTTDOTWLTWLVQY 302
Db 306 TIYQSWFSASNPANLPGLIWDHAGYLAOKDAPVIVGEFGKFPETISKQWNLTVGY 365
QY 303 LRPTAQYGADSFQWTFWSWNPDSGDTGILKDDQVTDVTKDGYLAPI 350
Db 366 LSSI---GISS---SPWAFENSGDTGIVKSDWVTPEQAKLDALAPI 407

RESULT 3
Q9AF65 PRELIMINARY; PRT; 727 AA.
ID Q9AF65
AC Q9AF65;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Cellulase CelA.
OS Clavibacter michiganensis (subsp. sepedonicus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococciaceae; Microbacteriaceae; Clavibacter.
OX NCBI_TaxID=31964;
RN [1]
RP SEQUENCE FROM N.A.
RA Lane M.; Haapalainen M.; Wahlroos T.; Kankare K.; Missinen R.;
RA Kassuwi S.; Mettler M.C.;
RT "The cellulase encoded by the native plasmid of Clavibacter
RT michiganensis subsp. sepedonicus plays a role in virulence and
RT contains an expansin-like domain.";
RL Physiol. Mol. Plant Pathol. 0:0-0(2001).
DR ENBL; AY007311; AAK16222.1; -.
DR HSSP; P54583; LECE.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001919; Bac celose-bind.
DR InterPro; IPR008965; Cellul bind.
DR InterPro; IPR007117; Expan_Lol_pi_C.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00553; CBM_2; 1.
DR Pfam; PF00150; cellulase; 1.
DR SMART; SM00637; CBD II; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
SQ SEQUENCE 727 AA; 75811 MW; 96DF9B664873985D CRC64;

Query Match 45.0%; Score 894.5; DB 2; Length 727;
Best Local Similarity 50.0%; Pred. No. 4.2e-61;
Matches 174; Conservative 48; Mismatches 115; Indels 11; Gaps 4;

QY 4 GYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDYRSMDOI KSLGYNTIRL 63
Db 49 GWLHTAGGKIIVTASGAPYIRGIAMFGMETSSCAPHGLDTITLGGMQHILKQMGFTTVRL 108
QY 64 PYSDILKPGTNPNSINFYQNMQDLOGLTSLQWMDKIVAYAGQIGLRIILDRHRPDCSGQ 123
Db 109 PFSNQCL-----AASGVTVGSADPSLAGLTPLOVMDHVVASAKSAGLDVILDQHRPDSGGQ 164
QY 124 SALWYTSVSEATWISDICALAQRKGNPTVGVFDLHNEPHDPACWCGDPSIDWRLAAE 183
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:30:29 / Search time 36.3333 Seconds  
(without alignments)  
3108.867 Million cell updates/sec

Title: US-09-997-504A-10  
Perfect score: 1987  
Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues 1017041

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 25.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertibrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1060	53.3	397	2 Q8RP23	Q8rp23 paenibacill
2	901.5	45.4	745	2 Q9K5C7	Q9k5c7 claviibacter
3	894.5	45.0	727	2 Q9AF65	Q9af65 claviibacter
4	800.5	40.3	518	2 Q9EV66	Q9evr6 xanthomonas
5	799.5	40.2	474	16 Q8P8S3	Q8pps3 xanthomonas
6	755	38.0	458	17 Q58925	Q58925 pyrococcus
7	690	34.7	482	16 Q97FK6	Q97fk6 clostridium
8	687.5	34.6	592	16 Q9PF60	Q9pf60 xylella fas
9	682.5	34.3	514	17 Q9V052	Q9v052 pyrococcus
10	649.5	32.7	614	16 Q8VAH4	Q8vah4 xylella fas
11	627.5	31.5	1000	2 Q24820	Q24820 thermophili
12	621	31.3	1426	2 Q9X3P6	Q9x3p6 caldicellul
13	620.5	31.2	660	2 Q9L3J2	Q9l3j2 clostridium
14	591.5	29.8	534	2 Q9EY00	Q9ey00 clostridium
15	582	29.3	616	2 Q7X2N2	Q7x2n2 thermomonas
16	513.5	25.8	425	2 Q93Q07	Q93q07 thermus cal

17	503	25.3	574	2 Q9S3V3	Q9s3v3 cellulomona
18	488	24.6	341	2 P96310	P96310 anaerocellu
19	403	20.3	630	2 Q8RJY7	Q8rjy7 stigmatella
20	328	16.5	565	16 Q9PF68	Q9pf68 xylella fas
21	315.5	15.9	569	16 Q87AG9	Q87ag9 xylella fas
22	276.5	13.9	590	16 Q8P513	Q8p513 xanthomonas
23	267	13.4	508	10 Q9LTM8	Q9ltm8 arabidopsis
24	258.5	13.0	522	10 Q9SAE6	Q9sae6 arabidopsis
25	221	11.1	526	10 Q9LFS2	Q9lfs2 arabidopsis
26	211.5	10.6	488	10 Q9LFE7	Q9lfe7 arabidopsis
27	196.5	9.9	555	10 Q7XUQ4	Q7xuq4 oryza sativ
28	190.5	9.6	354	16 Q8PEF2	Q8pef2 xanthomonas
29	184.5	9.3	561	10 Q9LTN0	Q9ltn0 arabidopsis
30	173.5	8.7	437	3 Q99036	Q99036 trichoderma
31	170	8.6	550	10 Q7XUQ5	Q7xuq5 oryza sativ
32	169.5	8.5	694	2 Q9Z187	Q9z187 bacillus st
33	169	8.5	439	3 Q9Z401	Q9z401 agaricus bi
34	168.5	8.5	377	3 Q00012	Q00012 aspergillus
35	168	8.5	439	3 Q9P893	Q9p893 agaricus bi
36	167	8.4	377	16 Q8PRD5	Q8prd5 xanthomonas
37	166.5	8.4	558	10 Q7XUQ6	Q7xuq6 oryza sativ
38	164.5	8.3	516	2 Q6L185	Q6l185 bacillus ci
39	160.5	8.1	357	16 Q8PRD3	Q8prd3 xanthomonas
40	160.5	8.1	722	16 Q97L56	Q97l56 clostridium
41	159.5	8.0	363	2 Q07652	Q07652 cellvibrio
42	157	7.9	364	2 Q9F0G8	Q9f0g8 rhizobium m
43	155.5	7.8	356	16 Q9PA12	Q9pa12 xylella fas
44	153.5	7.7	356	16 Q879X8	Q879x8 xylella fas
45	153.5	7.7	1449	10 Q8RU51	Q8ru51 oryza sativ

## ALIGNMENTS

## RESULT 1

Q8RP23 PRELIMINARY, PRT; 397 AA.

AC Q8RP23; DT 01-JUN-2002 (Tremblrel. 21, Created)

DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)

DE Endo-beta-1,4-glucanase

OS Paenibacillus sp. KRC8848P.

OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.

ON NCBI TaxID=109199;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-KCTC 8848P.

RA Park J.N., Kim H.O., Shin D.J., Lee H.B., Chun S.B., Bai S.;

RT "Cloning of Paenibacillus sp. endo-beta-1,4-glucanase gene and its co-

RT expression with Endomyces fibuliger beta-glucosidase gene in

RT Saccharomyces cerevisiae."

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF345984; AAL83749.1; "

DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR001547; Glyco\_hydro\_5.

DR Pfam; PF00150; cellulase; I.

DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.

SQ SEQUENCE 397 AA; 44539 MW; 1736A4C0C04E298F CRC64;

Query Match 53.3%; Score 1060; DB 2; Length 397;

Best Local Similarity 55.0%; Pred No. 2, 4e-74;

Matches 196; Conservative 44; Mismatches 102; Indels 16; Gaps 5;

QY 4 GYWHTSGREILDANNVPVRLAGINWFGPFCNYVYVHGLWSRDYRSMLDQIKSLGYNTRL 63

Db 37 GYHPTQKGNKVDEITGKEAFAFGLNWFLETPNTYTLHGLWSRSMDDMLDQVKEGYNTRL 96

QY 64 PYSDILKPGIMPNSINFEYQVNDQLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSQ 123

Db 97 PYSNQLFDSSRAUSDIDYK-NPDVLVGTPIQIMDKLIEKAGQKGIQLILDRHRPDCSQ 155

This Page Blank (uspto)



CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN  
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D90341; BAA14354.1; --  
 CC HSSP; P17901; LEDG. CBM 11.  
 CC InterPro; IPR005087; CBM 11.  
 CC InterPro; IPR002105; Dockerin\_1.  
 CC InterPro; IPR002048; EF-hand.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC Pfam; PF03425; CBM 11; 1.  
 CC Pfam; PF00150; cellulase; 1.  
 CC Pfam; PF00404; Dockerin\_1; 2.  
 CC PROSITE; PS00018; EF HAND; UNKNOWN 1.  
 CC PROSITE; PS00448; CLOS CELLULOSE RPT; 1.  
 CC PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.  
 CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 CC SIGNAL 1 24 POTENTIAL.  
 CC FT CHAIN 25 584 ENDOGLUCANASE D.  
 CC FT DOMAIN 25 328 CATALYTIC (BY SIMILARITY).  
 CC FT DOMAIN 329 353 PRO/THR-RICH (LINKER).  
 CC FT DOMAIN 354 584 CELLULOSE-BINDING (BY SIMILARITY).  
 CC FT DOMAIN 530 584 2 X 24 AA APPROXIMATE REPEATS.  
 CC FT REPEAT 530 552 1.  
 CC FT REPEAT 562 584 2.  
 CC FT ACT\_SITE 159 159 PROTON DONOR (BY SIMILARITY).  
 CC FT ACT\_SITE 264 264 NUCLEOPHILE (BY SIMILARITY).  
 CC FT ACT\_SITE 264 264  
 CC SEQUENCE 584 AA, 66061 MW, 0FC41257E81322C3 CRC64;  
 CC  
 CC Query Match 7.1%; Score 140.5; DB 1; Length 584;  
 CC Best Local Similarity 21.0%; Pred. No. 0.0018;  
 CC Matches 71; Conservative 52; Mismatches 130; Indels 85; Gaps 18;  
 CC  
 CC QY 21 VRIAGINWGFCTCNVHGLMSRDYRS-MLDQIKSLGNTIRLPYSDDILKPGTWPNISI 79  
 CC DB 32 VKRMGIGMNLGNTFDAPTEGSKAAQAYYDFDFKQAGFKHVRIPIRWDQHTLANSPTV 91  
 CC QY 80 NFYQMNDLQGLTSLQWMDKIVAYAGQIGLIRILDRHRPDCSGSALW-----YTSSVS-- 133  
 CC DB 92 DSNFLNR-----IETVIDMSLSRGFTVINSHD-----TWLMDNYSQNIGRF 134  
 CC QY 134 EATWISDLOALAQRYKNTVTGFDLHNEPHDPACWGCGDPSIDWRLAARAGNAVLSVN 193  
 CC DB 135 EKIW-----EQIAQRFGKSENLFETLNEPHG-----NITDSQIN--DMNKRILNIIRKTN 184  
 CC QY 194 PNLLIFVEGVQSYNGDSYWMGNLQAGQYVVLNVPNRLVTSADHYATSVGPQTWFGDP 253  
 CC DB 185 PIRNVI-----GAGYWNYSNSQLIP--NDPN-LIATFYH-----DP 222  
 CC QY 254 -TFPNMNGIWN-----KNGYLFQNIAPVWLGFEGFTLQSTDTQWLK 297  
 CC DB 223 YSFTHWQQTGTGKNDMDAIAVFNVKKWS---DKNIPVYLGEYG--VMGHSRDTSAV 277  
 CC QY 298 TLVQYLRPTA-QYGADSFQTFWSNWPNDSGDTGGILKD 334  
 CC DB 278 KWFDFVSDQAIHGFSGAW-----DNGVFGSDND 308

RESULT 14

GUNH\_CLOTM

ID GUNH\_CLOTM STANDARD; PRT; 900 AA.

AC P16218;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase H precursor (EC 3.2.1.4) (EGH) (Endo-1,4-beta-glucanase)  
 DE (Cellulase H).  
 GN CELH.  
 OS Clostridium thermoCELLUM.  
 CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 CC Clostridium.  
 CC NCBI\_TaxID=1515;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIB 10682;  
 RX MEDLINE=90323606; PubMed=2197182;  
 RA Yaquee E., Beguin P., Aubert J.-P.,  
 RT "Nucleotide sequence and deletion analysis of the cellulase-encoding  
 RT gene celH of Clostridium thermoCELLUM.";  
 RL Gene 89:61-67(1990).  
 CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-  
 CC GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
 CC GLUCANS.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS  
 CC WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN  
 CC MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO CELLULOSE FAMILY  
 CC A (FAMILY 5 OF GLYCOSYL HYDROLASES).  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION, BELONGS TO FAMILY 26 OF  
 CC GLYCOSYL HYDROLASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M31903; AAA23225.1; --  
 CC PIR; JH0157; JH0157.  
 CC HSSP; P07985; 1CEC.  
 CC InterPro; IPR005087; CBM 11.  
 CC InterPro; IPR002105; Dockerin\_1.  
 CC InterPro; IPR002048; EF-hand\_5.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC Pfam; PF03425; CBM 11; 1.  
 CC Pfam; PF00150; cellulase; 1.  
 CC Pfam; PF00404; Dockerin\_1; 2.  
 CC PROSITE; PS00018; EF HAND; UNKNOWN 1.  
 CC PROSITE; PS00448; CLOS CELLULOSE RPT; 2.  
 CC PROSITE; PS00659; GLYCOSYL HYDROL F5; 1.  
 CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 CC SIGNAL 1 44  
 CC FT CHAIN 45 900 ENDOGLUCANASE H.  
 CC FT DOMAIN 45 630 CATALYTIC (BY SIMILARITY).  
 CC FT DOMAIN 631 654 PRO/THR-RICH (LINKER).  
 CC FT DOMAIN 655 900 CELLULOSE-BINDING (BY SIMILARITY).  
 CC FT ACT\_SITE 460 460 PROTON DONOR (BY SIMILARITY).  
 CC FT ACT\_SITE 565 565 NUCLEOPHILE (BY SIMILARITY).  
 CC FT DOMAIN 833 895 2 X 24 AA APPROXIMATE REPEATS.  
 CC FT REPEAT 833 856 1.  
 CC FT REPEAT 872 895 2.  
 CC SEQUENCE 900 AA; 102415 MW; 973AFB1954FC346B CRC64;  
 CC  
 CC Query Match 6.9%; Score 137; DB 1; Length 900;  
 CC Best Local Similarity 20.2%; Pred. No. 0.0057;  
 CC Matches 70; Conservative 51; Mismatches 131; Indels 94; Gaps 18;

QY 40 GLWSDR-YRSMLDQIKSLGNTIRLPYSDDILKPGTWPNISINFYQMNDLQGLTSLQWMD 98

DB 352 GWSKSAEYVDFDFKAAQYKVRIPVRWBNHTMRTVPTID-----KAFLDRI 401

QY 99 KIVAYAGQIGLIRILDRHRPDCSGSALW-----YTSSVS--EATWISDLOALAQRYKGNP 152

Query Match 8.1%; Score 161.5; DB 1; Length 814;  
Best Local Similarity 21.8%; Pred. No. 6.3e-05;  
Matches 84; Conservative 59; Mismatches 132; Indels 111; Gaps 20;  
QY 6 WHTSGEILDANNVPVRIAG-----INWFGFETCNVYVHGLWS--RDYRSM 49  
DB 32 FQVSGCTKLLDASNELVMGRDISAIDLKVEIKLGNLNTLDAPTETAWGNPRITKX 91  
QY 50 LQDIKSLGYNITRLPYSDILKPGTMPSINFQYMNQDLQGLTSLOWMDKIYAYAGQIGL 109  
DB 92 IEKVRMGFNAVPVPTWD--THIGPAD--YKIDE-----AWLNRVEEVNVLDCOM 141  
QY 110 RIILDRHPCDSCQSALWYTSVSEATWISDLQA-----LAQYKGNPTV 154  
DB 142 YALINLHH-----DNTWIPYANQRSEKLVKVEQIATRFKDYDDH 185  
QY 155 VGFDLNHPDPACWCGDPSIDWRLAERAGNAVLSVNPENLLIFVEGVQSYNGDSYMWG 214  
DB 186 LLEFTWNEPRE-----VGSP-NEWMGGTVENRDVKNFN--LAVVNTIRAS-----G 229  
QY 215 GN-----LOGAGQYFVLN---VFN---RLVYSAHDY-----ATSV-GEQWTFSD- 252  
DB 230 GNNDKRFILVPTNAATGLDVALNDLVPNNDSRVIVSIHAYSPYFAMDVNGTSYWGSDY 289  
QY 253 --PTFNNMGIKNKNGYLFNQNIAFVWLGFEFTTQSTDTQTLKTLV---QYLREP 306  
DB 290 DKASLTSGLDLYNR-----FVNGRAVLIIGFG-----TIDKNLSRVAAEHYAREA 339  
QY 307 AQYGADSFQWTFWNPDSGDTGIL 332  
DB 340 VSRGIAVFWMDNGYVNPDAETVALL 365

RESULT 12  
GUNI\_RUMAL STANDARD; PRT; 406 AA.  
ID GUNI\_RUMAL  
AC P16216;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase I precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
DE (Cellulase) (EG-I).  
GN EG I.  
OS Ruminococcus albus.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae;  
OC Ruminococcus.  
OX NCBI\_TaxID=1264;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-58.  
RC STRAIN=F-40;  
RX MEDLINE=90078126; PubMed=2687251;  
RT Ohmura K., Kajino T., Kato A., Shimizu S.;  
RA "Structure of a Ruminococcus albus endo-1,4-beta-glucanase gene."  
RL J. Bacteriol. 171:6771-6775(1989).  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
CC hydrolases).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M30928; AAA26469.1; --  
DR PIR; A43722; A43722.  
DR HSRP; P17901; 1EDG.  
DR InterPro; IPR001547; Glyco\_hydro\_5.  
DR Pfam; PF00150; cellulase; 1.  
DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
DR

KW Cellulose degradation; Hydrolase; Glycosidase; Signal.  
FT SIGNAL 1 43  
FT CHAIN 44 406 ENDOGLUCANASE I.  
FT ACT\_SITE 210 210 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 330 330 NUCLEOPHILE (BY SIMILARITY).  
SQ SEQUENCE 406 AA; 45390 MW, 2E0172437B14FEA8 CRC64;  
Query Match 7.4%; Score 147.5; DB 1; Length 406;  
Best Local Similarity 22.9%; Pred. No. 0.00033;  
Matches 88; Conservative 49; Mismatches 152; Indels 95; Gaps 18;  
QY 15 DANNVPVRIAGIN-----WGFETCNVYVHGLWSR-----DYR 47  
DB 49 ETENVPVSQTHNDTMTVTSAKOLVAKMTNGWNLGNTMDTAQGLGSEVSWLPKVTNK 108  
QY 48 SMLDOIKSLGYNITRLPYSDILKPGTMPSINFQYMNQDLQGLTSLOWMDKIYAYAGQI 107  
DB 109 YNIDMLPEAGFNVLKIPVS-----WGNHIDDKYTSDFANMDRVQ---EIVNYGIDN 157  
QY 108 GLRIILDRHPCDSCQSALWYTSVSE-----ATWISDLQALQRYKGNPTVWVGP 158  
DB 158 GLYVILNTHHEE-----WTMPKPSKOGDIERIKAVWAQ---IADRFKGYDEHLIFE 206  
QY 159 LHNPHDPACWCGDPSIDWRLAERA-----GNVLSVNPENLLIFVEGVQSYNGDS 210  
DB 207 GLNEP-----RLRGEAEWTGTSEAREIINEYEKAFVETVRASGNGNDR 251  
QY 211 YWVGNGNLQAGQYF--VVLNVP---NRLVYSAHDYAT--SVGPQWTFSDPTFPNN--MPGI 262  
DB 252 CLMTCTYAASSAYNNLSAIELPEDSKLISVHAYLPVSFALDTKGTDKYDPEDTAIPEL 311  
QY 263 WNKWNGYLFNQNIAFVWLGFEFTTQSTDTQTLKTLVQYLRTAQYGADSFQWTFW 322  
DB 312 F-EHLNLFISKGIPVIVGFEFTGNTKENTEDR--VKCLEDYLAATAAKYDIPCVWMDNYA-R 368  
QY 323 PDSGDTGGILKD---DMQTVDTVK 343  
DB 369 IONGENFGEMRADLEWTFPDLIE 392

RESULT 13  
GUND\_CLOCE STANDARD; PRT; 584 AA.  
ID GUND\_CLOCE  
AC P25472;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase D)  
DE (Cellulase D) (EGCCD).  
GN CELCCD  
OS Clostridium cellulolyticum.  
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_TaxID=1521;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 35319;  
RX MEDLINE=92009193; PubMed=1916275;  
RA Shima S., Igarashi Y., Kodama T.;  
RT "Nucleotide sequence analysis of the endoglucanase-encoding gene,  
celCCD, of Clostridium cellulolyticum."  
RL Gene 104:33-38(1991).  
CC -!- FUNCTION: The biological conversion of cellulose to glucose  
CC generally requires three types of hydrolytic enzymes: (1)  
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)  
CC Exocellulohydrolases that cut the disaccharide cellobiose  
CC from the nonreducing end of the cellulose polymer chain; (3)  
CC Beta-1,4-glucosidases which hydrolyze the cellobiose and other  
CC short cello-oligosaccharides to glucose.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- PATHWAY: Cellulose degradation.  
CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS



RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AW;  
 RX MEDLINE=92138626; PubMed=1735723;  
 RA Huang J., Schell M.A.;  
 RT "Role of the two-component leader sequence and mature amino acid  
 sequences in extracellular export of endoglucanase EGL from  
 Pseudomonas solanacearum";  
 RL J. Bacteriol. 174:1314-1323 (1992).  
 RN [2]  
 RP SEQUENCE OF 1-112 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=89291722; PubMed=2738021;  
 RA Huang J., Sukordhaman M., Schell M.A.;  
 RT "Excretion of the egl gene product of Pseudomonas solanacearum";  
 RL J. Bacteriol. 171:3767-3774 (1989).  
 RN [3]  
 RP PROCESSING.  
 RX MEDLINE=90307678; PubMed=2195024;  
 RA Huang J., Schell M.A.;  
 RT "Evidence that extracellular export of the endoglucanase encoded by  
 egl of Pseudomonas solanacearum occurs by a two-step process  
 involving a lipoprotein intermediate";  
 RL J. Biol. Chem. 265:11628-11632 (1990).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 (Probable).  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 hydrolases).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M84922; AAA61980.1; -  
 CC PIR; A42649; A42649.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC InterPro; IPR000437; Prok\_LipoProt\_S.  
 CC Pfam; PF00150; cellulase\_1.  
 CC PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 CC PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 CC Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;  
 KW Lipoprotein; Signal; Palmitate.  
 FT SIGNAL 1 19  
 FT PROPEP 20 45  
 FT CHAIN 46 426  
 FT LIPID 20 20 N-palmitoyl cysteine.  
 FT LIPID 20 20 S-diacylglycerol cysteine.  
 FT ACT SITE 249 249 PROTON DONOR (BY SIMILARITY).  
 FT ACT SITE 361 361 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 426 AA; 45578 MW; 51E13AD442CF4A8 CRC64;  
 Query Match 8.8%; Score 175; DB 1; Length 426;  
 Best Local Similarity 23.7%; Pred. No. 2.6e-06;  
 Matches 84; Conservative 55; Mismatches 135; Indels 80; Gaps 19;  
 QY 25 GINWFGFCYCNVYVHGLWSDY-----RSMIDQIKSLGYNTIRLPYSDDILKPGTMPNSIN 80  
 Db 121 GVSLAGAERGESLPCTGYSNVIYPSADSVTYVKNKGNLVLPRFWRQLP-----172  
 QY 81 FVQMNO--DLQGLTSLOVNDKIVAVAGQICLIILDRHPCDCSGSALWY-----TSSVS 133  
 Db 173 --TLNQFFDANELSRITGTFVNAVATGQ---TVLLDPH-----NYARYGNVIGSSAVP 221  
 QY 134 EATWISDQLAQRKYKNGTIVVGFDLHNEFHPACWCGDPSIDWRLAAERAGNAVLSVN 193  
 Db 222 NSAYADFWRRLATQFKSNRVI-LGLMNEPNSN-----PTEQWLSGANAEIAAIRSAN 273  
 QY 194 PNLLIFVEGVQSINGSYWGGNMQAGQYFV--LNPV-NRLIVYSAHDY-----240

Db 274 ASNVFVFG-NAWTG-AHSWNQWYTPNGTVMKGNDFVHQLVDFVHQVLDGSSGQSA 331  
 QY 241 ---ATSVGPQTWFSDDPTFPNNMPGINKKNGWYLFNQNIAPIVWLGEGTTLQSTDTWLK 297  
 Db 332 NCVSATIGAQR-LQDFT-----TWLRNGY-----RGLGEGFAASNDTCNQA-VS 375  
 QY 298 TLVOYLRPTAQYGADSFQWTFWSNPDSGDTGGILK--DDWQTVDTVKDGYLAP 349  
 Db 376 NMLTFVKNADV-----WTGNWAGGPPWGGWYISIEPSNGVDKPFQMSVLAP 423  
 RESULT 11  
 ID\_GUNE\_CLOTM STANDARD; PRT; 814 AA.  
 AC P10477;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Endoglucanase E precursor (EC 3.2.1.4) (EGE) (Endo-1,4-beta-glucanase)  
 DE (Cellulase E).  
 GN CELE  
 OS Clostridium thermocellum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OX NCBI\_TaxID=1515;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-58.  
 RX MEDLINE=9917992; PubMed=3066698;  
 RA Hall J., Hazlewood G.P., Barker P.J., Gilbert H.J.;  
 RT "Conserved reiterated domains in Clostridium thermocellum  
 endoglucanases are not essential for catalytic activity";  
 RL Gene 69:29-38 (1988).  
 CC -!- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-  
 GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-  
 GLUCANS.  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- DOMAIN: A 24 RESIDUES DOMAIN IS REPEATED TWICE IN THIS ENZYME AS  
 WELL AS IN OTHER C.THERMOCELLUM CELLULOSE ENZYMES. THIS DOMAIN  
 MAY FUNCTION AS THE BINDING LIGAND FOR THE SL COMPONENT.  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 hydrolases).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M22759; AAA23224.1; -  
 CC PIR; JTO347; CZCLEM.  
 CC HSSP; P17901; 1EDG.  
 CC InterPro; IPR002105; Dockerin\_1.  
 CC InterPro; IPR001547; Glyco\_hydro\_5.  
 CC InterPro; IPR001087; Lipase\_GDSL.  
 CC Pfam; PF00150; cellulase; 1.  
 CC Pfam; PF00404; Dockerin\_1; 2.  
 CC PROSITE; PS00657; Lipase\_GDSL; 1.  
 CC PROSITE; PS00448; CLOS\_CELLULOSE\_RPT; 2.  
 CC PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 CC Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 KW SIGNAL 1 34  
 FT CHAIN 35 814 ENDOGLUCANASE E.  
 FT ACT SITE 193 193 PROTON DONOR (BY SIMILARITY).  
 FT ACT SITE 316 316 NUCLEOPHILE (BY SIMILARITY).  
 FT DOMAIN 415 474 2 X 24 AA APPROXIMATE REPEATS.  
 FT REPEAT 415 438 1.  
 FT REPEAT 451 474 2.  
 SQ SEQUENCE 814 AA; 90244 MW; C6FA24B8D1523632 CRC64;

DOMAIN 281 748 CATALYTIC.  
 FT DISULFID 39 133 BY SIMILARITY.  
 FT ACT SITE 503 PROTON DONOR (BY SIMILARITY).  
 FT ACT\_SITE 653 NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 748 AA; 80197 MW; AC96104137932B76 CRC64;  
 Query Match 14.7%; Score 291.5; DB 1; Length 748;  
 Best Local Similarity 23.4%; Pred. No. 4.5e-15;  
 Matches 111; Conservative 61; Mismatches 138; Indels 165; Gaps 23;  
 7 HTSGREILDANNVPVRIAGINWFGFTCN-----TVVHGLW-----S 43  
 Db 290 NTNGNLTGDKQLLPARG--NWFLGRHPSNDADNPSGAPVELYAGNWWNNSSQSGS 347  
 Qy 44 RYRSMLDQIKSLGYNTIRLPYDDIL-----KPTMPNSINFYQMDLQGLTSLQVMDK 99  
 Db 348 RTIQQTWTELKQOQITWLRPIAPQTLNDNDPQGRSPNLKN-----HQSIQSNARQALD 403  
 Qy 100 IVAVAGQIGRLIILDRH-----RPDCSGQSALWY-----TSSVS 133  
 Db 404 FIKLADQNDIQIFIDHSCSNVYWGVRAGRLDARPPYVDNRVGYDTREYSCSATNPS 463  
 Qy 134 EAT-----WISDLQALA--ORYKGNPTVVGFDLHNEPHDPACGCGDPSIDWRLAA 182  
 Db 464 SVTRIHAYDKQKLANLREIAGLSAKLGVSNLIGIDVFNEFYD-YTWA-----EWKGMV 516  
 Qy 183 ERAGNAVLSVNPMLLIPVEGVQSVNGDSY-----WWGNGNLQAGQY 223  
 Db 517 EEAQYAINNEVNPMLLIVEGI-SANANTQDGTPTSPVPHGSTDLPNPNNGENLYEAGAN 575  
 Qy 224 PVVLNVP-NRLVYSAHDYATSVGPOTWFSPT-----FPNNMFG 261  
 Db 576 PP--NIPKRLRLSPHYTGSVFRQFMDPAQTECAGLEGDEAAQACRIVINPTVLEQ 633  
 Qy 262 IWKNNGYLFNQNIAPVWLGEFG-----TTLQSTTDQWTKTLVQYL 303  
 Db 634 GWSEHFGYRELGYG-ILIEFGGNDWPGAKSSQADRNAWSHTTNDVQOQQAASYF 692  
 Qy 304 RPTAQYQADSFQWTFWNPDSGDTGILKDDQTVDTVKDGLYLAPIKSIFFPV 358  
 Db 693 K---RKGINA---CYWSNPEASDTMG-----W-----YLPD-----WDPV 722  
 RESULT 9  
 ID\_GUN1\_RALSO STANDARD; PRT; 424 AA.  
 AC P58599;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase)  
 GN EGI OR RSF0162 OR RS05516.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OG Plasmid megaplasmid.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502 (2002).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor

(Probable).  
 -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl hydrolases).  
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL; AL646076; CAD17313.1; ALT INIT.  
 InterPro; IPR001547; Glyco\_hydro\_5.  
 InterPro; IPR000437; Prok\_Lipprot\_S.  
 Pfam; PF00150; cellulase; 1.  
 PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 DR Cellulose degradation; Hydrolase; Glycosidase; Zymogen; Membrane;  
 KW Lipoprotein; Signal; Plasmid; Complete proteome; Palmitate.  
 FT SIGNAL 1 19 BY SIMILARITY.  
 FT PROPEP 20 43 BY SIMILARITY.  
 FT CHAIN 44 424 ENDOGLUCANASE.  
 FT LIPID 20 20 S-diacylglycerol cysteine (By similarity).  
 FT ACT\_SITE 247 247 N-palmitoyl cysteine (By similarity).  
 FT ACT\_SITE 359 359 PROTON DONOR (BY SIMILARITY).  
 FT NUCLEOPHILE (BY SIMILARITY).  
 SQ SEQUENCE 424 AA; 45080 MW; 5BF0D161AB7C87EF CRC64;  
 Query Match 9.4%; Score 187; DB 1; Length 424;  
 Best Local Similarity 25.0%; Pred. No. 3e-07;  
 Matches 85; Conservative 51; Mismatches 122; Indels 82; Gaps 19;  
 Qy 35 NYVHGLRSYRSMDOIKSGYNTIRLPYSDILKPGTNPNSINFYQNMQ--DLQGLT 92  
 Db 139 NYIYPSASATY-----YKNGMNLVLPFRWELQP-----TLNQALDANELS 182  
 Qy 93 SLQVMDKIVAVAGQIGRLIILDRHPPDCSGQSALWY-----TSSVSEATWISDIQALAQ 147  
 Db 183 RLTFGVNAVTAAGQ--TVLLDPH-----NYARYYGNVIGSSAVPNSAYADFRRVATQ 233  
 Qy 148 YKGNPTVVGFDLHNEPHDPACGCGDPSIDWRLAERAGNAVLSVNPMLLIPVEGVQSYN 207  
 Db 234 FKGNARV-FGLMNEPNSM-----PTEQWLSGANALAAAIRSANASNVVFPVFG-NAWT 284  
 Qy 208 GDSYVWNGNLQAGQYVYV--LVNPNR-LVYSAHDY-----ATSVGPQTWFS 251  
 Db 285 G-AHSWNQNWYGTPTNGTVMKGNDEGRNLVFEVHQYLDGSSGQSASCVSATIGAERL-- 341  
 Qy 252 DPTFPNNMFGIWNKNGYLFNQNIAPVWLGEFGITLQSTTDQWTKTLVQYLRTAQYGA 311  
 Db 342 -QDFTN-----WLRNGY-----RGFLGFGAASNDTCQA-VANMLTFVKNNAADV-- 385  
 Qy 312 DSFQWTFWSNPNDSGDTGILK--DDWQTVDTVKDGYLAP 349  
 Db 386 ----WTGWAWAGGFWPGWGYMYSIEPNSGVNDKPMQSVLAP 421  
 RESULT 10  
 ID\_GUN2\_RALSO STANDARD; PRT; 426 AA.  
 AC P17974;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN EGI.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Ralstonia.  
 OX NCBI\_TaxID=305;



DR Pfam: PF00404; Dockerin 1; 2.  
 DR PROSITE; PS00018; EF HAND; UNKNOWN 2.  
 DR PROSITE; PS00448; CLOS CELLULOSE\_RPT; 2.  
 DR PROSITE; PS00559; GLYCOSYL HYDROL\_F5; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 566  
 FT ACT\_SITE 226 226  
 FT ACT\_SITE 381 381  
 FT DOMAIN 503 549  
 FT REPEAT 503 526  
 FT REPEAT 536 549  
 SQ SEQUENCE 566 AA; 63199 MW; 2CC9DE1AD87C3178 CRC64;  
 Query Match 30.28; Score 600.5; DB 1; Length 566;  
 Best Local Similarity 34.28; Pred. No. 3.2e-39;  
 Matches 139; Conservative 55; Mismatches 143; Indels 69; Gaps 15;  
 QY 6 W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMQLDQIKSLGYNTIRLP 64  
 DB 44 WLHCKGNKIYDMYGNVWLGTANWFGFENCFHGAW-YDVKTILTSIADRGINLLRIP 102  
 QY 65 YSDIL-----KPG-----TPNSINFYQNDLQ-----LTSLOWMDKIVAYAGQI 107  
 DB 103 ISTEILYSMMGKPNPVSSTASNNPPYHVNPFDYDPETDDVKNSEIFDIIMGYCKEL 162  
 QY 108 GLRIILDRHPDC--SQS-ALWY---TSS---VSEATWISDLQALAQRYKGNPTVVGF 158  
 DB 163 GIKWIDIHSDPANNNGHNYELVYKETSCTGVTTQWIDTLWLADRYKNDITIAED 222  
 QY 159 LHNEPDP-----ACWCGDPSIDWRLAERAGNAVLSVNPILLIFVEGVQSY- 206  
 DB 223 LKNEPHGRGYTAEPVPLKAKNDSTDENWKAATCAKAILVNPVKVLIVTEGVEQVP 282  
 QY 207 -----NGD-----SYWNGNLCAGQYVVLNVPN-RLVYSADHYATS 244  
 DB 283 KTEGYTYDTPDINGATGDSAPYSAWGNLGRVKDYPIDGLPNSQIVYSPHDGSPV 342  
 QY 245 GPQTFP-SDTPFNMPGINKWNGYLFNQNIAPVWLGEFTGLTQSTTDQTKLTLVQYL 303  
 DB 343 YAPWFEKFTMQTLDDYWDYTWAYTHDQGIAPILIGENGHMGDKGNKQWTLRLDYI 402  
 QY 304 RPTAQGADSFQTFWGNPDSGDTGILKDDQWTVTKDGLAP 349  
 DB 403 -----VONRIHTFWCINPNSGDTGGLGNDWSTWDEAKYALLKP 442  
 RESULT 6  
 GUND\_CELFI  
 ID GUND\_CELFI STANDARD; PRT; 747 AA.  
 AC PS0400;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)  
 DE (Cellulase).  
 GN CEND.  
 OS Cellulomonas fimi.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Micrococineae; Cellulomonadaceae; Cellulomonas.  
 OC NCBI\_TaxID=1708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-93209933; PubMed=8458833;  
 RA Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.;  
 RT "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase  
 D (Cend), a family A beta-1,4-glucanase."  
 RL J. Bacteriol. 175:1910-1918(1993).  
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
 CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
 CC -!- PATHWAY: Cellulose degradation.  
 CC -!- SIMILARITY: Contains 2 fibronectin type III domains.  
 CC -!- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)

CC domain.  
 CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
 CC hydrolases).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L02544; AAA23089.1; -.  
 DR HSP; P07986; 1EXG.  
 DR InterPro; IPR001919; Bac celose-bind.  
 DR InterPro; IPR008965; Cellul bind.  
 DR InterPro; IPR008967; FN-III-like.  
 DR InterPro; IPR003561; FN-III.  
 DR InterPro; IPR001547; Glyco\_hydro\_5.  
 DR Pfam; PF00553; CBM 2; 1.  
 DR Pfam; PF00041; fn3; 2.  
 DR SMART; SM00637; CBD II; 1.  
 DR SMART; SM00060; FN3\_2.  
 DR PROSITE; PS00659; GLYCOSYL\_HYDROL\_F5; 1.  
 KW Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.  
 FT SIGNAL 1 39  
 FT CHAIN 40 747  
 FT DOMAIN 448 542  
 FT DOMAIN 546 639  
 FT ACT\_SITE 208 208  
 FT ACT\_SITE 349 349  
 SQ SEQUENCE 747 AA; 78936 MW; BD15473C9D842BD CRC64;  
 Query Match 27.7%; Score 549.5; DB 1; Length 747;  
 Best Local Similarity 35.0%; Pred. No. 4.1e-35;  
 Matches 132; Conservative 58; Mismatches 142; Indels 45; Gaps 15;  
 QY 1 AGGGVW-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMQLDQIKSLGYN 59  
 DB 40 ATGDDWLHVEGNTIVDSTGKEAILSGVNWFGFNASERVFHGLWSGNITQITQQMAQRGIN 99  
 QY 60 TRLPYSDIL--KPGTW--PNSINFYQNDLQGLTSLQWMDKIVAYAGQIGRLILD 114  
 DB 100 VVRVSTQLLEWAGTFLEPN-VNTY-ANPELEGNLSLQIFEWLTLCKYGIKVFLD 157  
 QY 115 RH--RPDCSGQ-SALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPH-----D 165  
 DB 158 VHSADNSGHVYNNWKGDTTDDVYEGWEWAATRWKDDDTIVGADIKNEPHGTQGST 217  
 QY 166 PACWCGDPSIDWRLAERAGNAVLSVNPILLIFVEGVQSYN-----GDSY--W 212  
 DB 218 RAKWDGTTDKDNFKHFAETASKKILAINPNWLVFVEGVETLTPKGVPTWSTGLTDYGTW 277  
 QY 213 WGNLQAGQYVVLNV--PNELVYSADHYATSVDGQTFW-SDTPFNMPGINKWNGYL 270  
 DB 278 WGNLGRVDRDHPIDLAGHODQLVYSPHDYGLVFDQKWFQKDFDKASLTADVWGNPLFI 337  
 QY 271 FNQNTAPVWLGEFGTTL-QSTTDQW---LKTLYQLRPTAQYGADSPQTFWNSNPDSG 326  
 DB 338 HDIEDIAPLLIGEWGRGLQDFRQDKWMAALRDLVAERR-----LSQTFVWLNPNSG 386  
 QY 327 DTGGLKDDQWQTVTK 343  
 DB 369 DTGGLLDDWKTWDEYK 405  
 RESULT 7  
 GUND\_CLOTM  
 ID GUND\_CLOTM STANDARD; PRT; 563 AA.  
 AC P04956;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)





FT	DOMAIN	401	PRO/SER/THR-RICH (LINKER).
FT	DOMAIN	462	CELLULOSE-BINDING (BY SIMILARITY).
FT	ACT_SITE	203	PROTON DONOR.
FT	ACT_SITE	323	NUCLEOPHILE.
FT	DISULFID	75	
FT	DISULFID	209	
FT	STRAND	48	
FT	TURN	51	
FT	STRAND	53	
FT	TURN	57	
FT	STRAND	61	
FT	STRAND	65	
FT	STRAND	72	
FT	TURN	73	
FT	TURN	76	
FT	TURN	80	
FT	HELIX	87	
FT	TURN	97	
FT	STRAND	101	
FT	HELIX	108	
FT	TURN	113	
FT	HELIX	127	
FT	TURN	130	
FT	HELIX	134	
FT	TURN	148	
FT	STRAND	151	
FT	STRAND	160	
FT	TURN	161	
FT	STRAND	163	
FT	HELIX	175	
FT	TURN	189	
FT	TURN	193	
FT	STRAND	195	
FT	TURN	206	
FT	STRAND	209	
FT	TURN	215	
FT	STRAND	218	
FT	HELIX	219	
FT	TURN	231	
FT	TURN	235	
FT	STRAND	238	
FT	STRAND	244	
FT	STRAND	246	
FT	TURN	248	
FT	STRAND	250	
FT	TURN	254	
FT	STRAND	257	
FT	TURN	259	
FT	STRAND	274	
FT	TURN	283	
FT	HELIX	289	
FT	TURN	294	
FT	HELIX	300	
FT	TURN	308	
FT	HELIX	309	
FT	TURN	314	
FT	STRAND	319	
FT	HELIX	331	
FT	TURN	344	
FT	HELIX	347	
FT	TURN	351	
FT	STRAND	356	
FT	TURN	367	
FT	STRAND	373	
FT	TURN	375	
FT	STRAND	380	
FT	HELIX	382	
FT	TURN	388	
FT	HELIX	389	
SQ	SEQUENCE	562 AA;	60747 MW; 84E6256406A35041 CRC64;
Query Match 99.5%; Score 1978; DB 1; Length 562;			
Best Local Similarity 99.7%; Pred. No. 3e-146;			
Matches 194; Conservative 43; Mismatches 107; Indels 12; Gaps 4;			
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	AGGYYWHTSGREILDANNVPRVIRAGINWFGFETCNVYVHGLMSRDYRSMLDQIKSLGYNT	60
Db	42	AGGYYWHTSGREILDANNVPRVIRAGINWFGFETCNVYVHGLMSRDYRSMLDQIKSLGYNT	101
Qy	61	IRLPYSDDILKPGTMPNSINFYQNDLQGLTSLOWMDKIVAYAGQIGRIILLDRHPDC	120
Db	102	IRLPYSDDILKPGTMPNSINFYQNDLQGLTSLOWMDKIVAYAGQIGRIILLDRHPDC	161
Qy	121	SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGSPSIDWEL	180
Db	162	SGQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGSPSIDWEL	221
Qy	181	AAERAGNAVLSVNPILLIFVEGVOSYNGDSYWGNGNLQAGQYVVLNPNRNVLYSAHDY	240
Db	222	AAERAGNAVLSVNPILLIFVEGVOSYNGDSYWGNGNLQAGQYVVLNPNRNVLYSAHDY	281
Qy	241	ATSVGPOTWFSDDPTFPNNMPCINWKNWGYLFNQNIAPVWMLGEFGTTLQSTTDQTLKTLV	300
Db	282	ATSVGPOTWFSDDPTFPNNMPCINWKNWGYLFNQNIAPVWMLGEFGTTLQSTTDQTLKTLV	341
Qy	301	QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDQVTVTVKDGXHLAPIKKSIFDPV	358
Db	342	QYLRPTAQYGADSFQWTFWSNPDSDGTGILKDDQVTVTVKDGXHLAPIKKSIFDPV	399
RESULT 2			
GUN_PAEPO STANDARD; PRT; 397 AA.			
ID	GUN_PAEPO	AC	P23548;
DT	01-NOV-1991 (Rel. 20, Created)	DT	01-FEB-1996 (Rel. 33, Last sequence update)
DE	28-FEB-2003 (Rel. 41, Last annotation update)	DE	Endoglucanase (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).
OS	Paenibacillus polymyxa (Bacillus polymyxa).	OS	Paenibacillus polymyxa (Bacillus polymyxa).
OC	Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.	OC	Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Paenibacillus.
OX	NCBI_TaxID=1406;	OX	NCBI_TaxID=1406;
RN	[1]	RN	[1]
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RX	MEDLINE=90170877; PubMed=2307659;	RX	MEDLINE=90170877; PubMed=2307659;
RA	Baird S.D., Johnson D.A., Seligy V.L.;	RA	Baird S.D., Johnson D.A., Seligy V.L.;
RT	"Molecular cloning, expression, and characterization of	RT	"Molecular cloning, expression, and characterization of
RT	endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus	RT	endo-beta-1,4-glucanase genes from Bacillus polymyxa and Bacillus
RT	circulans."	RT	circulans."
RL	J. Bacteriol. 172:1576-1586(1990).	RL	J. Bacteriol. 172:1576-1586(1990).
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic	CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC	linkages in cellulose, lichenin and cereal beta-D-glucans.	CC	linkages in cellulose, lichenin and cereal beta-D-glucans.
CC	-!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl	CC	-!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl
CC	hydrolases).	CC	hydrolases).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its	CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way	CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial	CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@sb-sib.ch).	CC	or send an email to license@sb-sib.ch).
CC	EMBL; M33791; AAA22631.1; -.	CC	EMBL; M33791; AAA22631.1; -.
DR	PIR; A35136; A35136.	DR	PIR; A35136; A35136.
DR	HSSP; P54583; 1RCE.	DR	HSSP; P54583; 1RCE.
DR	InterPro; IPR001547; Glyco_hydro_5.	DR	InterPro; IPR001547; Glyco_hydro_5.
DR	Fram; PF00150; cellulase; I.	DR	Fram; PF00150; cellulase; I.
DR	PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.	DR	PROSITE; PS00659; GLYCOSYL HYDROL_F5; 1.
KW	Cellulose degradation; Hydrolase; Glycosidase.	KW	Cellulose degradation; Hydrolase; Glycosidase.
FT	ACT_SITE	194	PROTON DONOR (BY SIMILARITY).
FT	ACT_SITE	317	NUCLEOPHILE (BY SIMILARITY).
SQ	SEQUENCE	397 AA;	44357 MW; B9C2E802C04F0A2A CRC64;
Query Match 52.9%; Score 1052; DB 1; Length 397;			
Best Local Similarity 54.5%; Pred. No. 1.7e-74;			
Matches 194; Conservative 43; Mismatches 107; Indels 12; Gaps 4;			

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM protein: - protein search, using sw model

Run on: August 2, 2004, 16:29:59 ; Search time 8.3333 Seconds  
(without alignments)

2236.934 Million cell updates/sec

Title: US-09-997-504A-10

Perfect score: 1987

Sequence: 1 AGGGYWHTSREILDANNVP.....VDTKDYLIPIKSSIFDPV 358

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1978	99.5	562	GUN1_ACICE	P54583 acidothermu
2	1052	52.9	397	GUN_PABPO	P23548 paenibacill
3	800.5	40.3	484	GUN_XANCP	P19487 xanthomonas
4	824.5	31.4	1039	GUNB_CALSA	P10474 c endogluca
5	600.5	30.2	566	GUNB_CLOTM	Q05352 clostridium
6	549.5	27.7	747	GUNB_CELFI	P50400 cellulomona
7	522	26.3	563	GUNB_CLOTM	P04956 clostridium
8	291.5	14.7	748	GUNB_PSEFL	P27033 pseudomonas
9	187	9.4	424	GUN1_RALSO	P58599 ralstonia s
10	175	8.8	426	GUN2_FALSO	P17974 ralstonia s
11	161.5	8.1	814	GUNE_CLOTM	P10477 clostridium
12	147.5	7.4	406	GUN1_RUMAL	P16216 ruminococcu
13	140.5	7.1	584	GUN2_CLOCE	P25472 clostridium
14	137	6.9	900	GUNH_CLOTM	P16218 clostridium
15	136.5	6.9	388	GUN3_HUMIN	Q12624 humicola in
16	136	6.8	409	GUNB_RUMAL	P23661 ruminococcu
17	132	6.6	438	EXG_CANAL	P29717 candida alb
18	128	6.4	448	GUN_CLOSA	P15704 clostridium
19	126	6.3	825	GUN3_BACSA	P19570 bacillus sp
20	122.5	6.2	517	GUNA_CLOLO	P54937 clostridium
21	120.5	6.1	658	GUN3_FIBSU	P14250 fibrobacter
22	120	6.0	336	GUNA_RUMFL	P16169 ruminococcu
23	120	6.0	440	GUNA_CLOCL	P28621 clostridium
24	119	6.0	445	SPR1_YEAST	P32603 saccharomyc
25	117.5	5.9	505	GUNV_ERWCA	Q47096 erwania car
26	117	5.9	501	YBQ6_YEAST	P38081 saccharomyc
27	115.5	5.8	504	GUNW_ERWCA	Q59395 erwania car
28	113.5	5.7	418	GUN2_TRIRE	P07982 trichoderma
29	113	5.7	444	GUNN_ERWCA	Q59394 erwania car
30	112.5	5.7	429	GUNA_BUTFI	P22541 butyrivibri
31	112.5	5.7	532	HEXA_DICDI	P13723 dictyosteli
32	112	5.6	764	YIS7_YEAST	P40566 saccharomyc
33	111	5.6	562	EXG2_YEAST	P52911 saccharomyc

34 110 5.5 515 1 GUND\_CLOCL  
35 109.5 5.5 475 1 GUNA\_CLOCE  
36 108.5 5.5 459 1 GUNA\_STRLI  
37 108.5 5.5 466 1 GUN5\_THRFU  
38 107.5 5.4 800 1 GUN\_BACSI  
39 107 5.4 800 1 GUN4\_RUMAL  
40 106.5 5.4 499 1 GUN2\_BACSU  
41 105.5 5.3 522 1 MAL2\_DRONE  
42 105.5 5.3 547 1 GUN1\_BUTFI  
43 104.5 5.3 364 1 GUNA\_RUMAL  
44 104.5 5.3 499 1 GUN3\_BACSU  
45 104 5.2 1024 1 SYIP\_STRAU

P28623 clostridium  
P17901 clostridium  
P27035 streptomyc  
Q01786 thermomonas  
P06564 bacillus sp  
Q07940 ruminococcu  
P10475 bacillus su  
P07190 drosophila  
P20847 butyrivibri  
P23660 ruminococcu  
P23549 bacillus su  
P41368 staphylococ

#### ALIGNMENTS

RESULT 1  
ID GUN1\_ACICE STANDARD; PRT; 562 AA.  
AC P54583;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Endoglucanase E1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase E1)  
DE (Cellulase E1) (Endocellulase E1).  
OS Acidothermus cellulolyticus.  
OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;  
OC Frankineae; Acidothermaceae; Acidothermus.  
OX NCBI\_TaxID=28049;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 43068 / 11b;  
RA Layman R.A., Himmel M.E., Thomas S.R.;  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 42-398.  
RX MEDLINE=96346058; PubMed=8718854;  
RA Sakon J., Adney W.S., Himmel M.E., Thomas S.R. P.A.;  
RT "Crystal structure of thermostable family 5 endocellulase E1 from  
RT Acidothermus cellulolyticus in complex with cellotetraose.";  
RL Biochemistry 35:10648-10660(1996).  
CC -!- FUNCTION: Has a very high specific activity on  
CC carboxymethylcellulose.  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
CC linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC -!- MISCELLANEOUS: Thermostable enzyme with an optimal temperature of  
CC 81 degrees Celsius.  
CC -!- SIMILARITY: Belongs to cellulase family A (family 5 of glycosyl  
CC hydrolases).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U33212; AAA75477.1; -  
CC PDB; 1ECE; 14-OCT-96.  
CC PDB; 1COD; 23-JUL-99.  
CC InterPro; IPR001919; Bac\_celose-bind.  
CC InterPro; IPR008965; Cellul bind.  
CC InterPro; IPR001547; Glyco\_hydro\_5.  
CC InterPro; IPR00553; CBM 2; 1.  
CC Pfam; PF00150; cellulase; 1.  
CC SMART; SM00637; CBD\_II; 1.  
CC PROSITE; P500659; GLYCOSYL\_HYDROL\_F5; 1.  
CC Cellulose degradation; Hydrolase; Glycosidase; Signal; 3D-structure.  
CC SIGNAL 1 41  
CC CHAIN 42 562  
CC DOMAIN 42 400  
CC CATALYTIC. ENDOGLUCANASE E1.



This Page Blank (uspto)

QY	68	DILKPGTSPNSINFIYQ-----NNQDLQG-----LTSLQVMDKIYVAGQIGLRI	111	QY	330	-GILKDDWQTV	339
Db	66	DLMTNETLANNVTVRQSFQSLGLNDDIVGFQTNPNPSIIDLPLIEAYKTVVTVTLGNNDVMV	125	Db	365	YGMLDANWHNV	375
QY	112	ILDRH--RPD--CSGQ--SALWTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD	165	Search completed: August 2, 2004, 16:37:37			
Db	126	ILDNHLTKPGWCCANDGNGFFGQDFPTVVAALKQMAATFNGVSNVVGMSLRNELRG	185	Job time : 13.3333 secs			
QY	166	PACWGCDDPSIDWRLAARAGNAVLSVNPENLIFVEGVSYNGDSYWWGNGLQAGQYFV	225				
Db	186	PK-----QNVNDWFKYMQQGAEAHSAANKVILSGL-SFDAD-----LSFVSRPV	232				
QY	226	VLVNPNRLVYSADHYATSVGFQWFSDFPTFPNNMPG-IWNK---NWGYLPNQNIAPVWL	281				
Db	233	KLSTGKLVFELHWYSFSDG-NSWAANN--PNDICGRVLNRI GNGGGYLLNQGF-PLFLS	288				
QY	282	EFGTTLQ--STTQDTLKTIVQLRPTAQXGADSFQWTFW-----SNPDSGDTG-----G	330				
Db	289	EFGIDERGVTNDRYFGCL-----TGWAENDVDWLSWALTCSYYLRQKYGKMEYYG	342				
QY	331	ILKDDWQTV	339				
Db	343	VLDSDWISV	351				
RESULT 15							
T51476							
hypothetical protein K3M16_70 - Arabidopsis thaliana							
C:Species: Arabidopsis thaliana (mouse-ear cress)							
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000							
C:Accession: T51476							
R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew							
submitted to the Protein Sequence Database, August 2000							
A:Reference number: Z25394							
A:Accession: T51476							
A:Status: preliminary							
A:Molecule type: DNA							
A:Residues: 1-526 <SAT>							
A:Cross-references: EMBL:AL391150							
A:Experimental source: cultivar Columbia; BAC clone K3M16							
C:Genetics:							
A:Map position: 5							
A:Introns: 135/3; 218/2; 327/2; 393/1							
A:Note: K3M16_70							
Query Match 11.1%; Score 221; DB 2; Length 526;							
Best Local Similarity 23.5%; Pred. No. 1.6e-09;							
Matches 87; Conservative 60; Mismatches 160; Indels 64; Gaps 14;							
QY	8	TSGREILDANNVPVRIAGINWFGPETCNVYVHGLWSRDYRSMLDQIKSLGYNTRILPY--	65	QY	8	TSGREILDANNVPVRIAGINWFGPETCNVYVHGLWSRDYRSMLDQIKSLGYNTRILPY--	65
Db	30	TKSRWIVNNKCHRVKLACANW-PSHLKPVVAEGLSSQPMDSISKIKDMGFNCVRLTWPL	88	Db	30	TKSRWIVNNKCHRVKLACANW-PSHLKPVVAEGLSSQPMDSISKIKDMGFNCVRLTWPL	88
QY	66	-----SDDLKPGTSPNSINFIYQNNQDLQGLTS-----LQVMDKIYVAGQIGLRI	111	QY	66	-----SDDLKPGTSPNSINFIYQNNQDLQGLTS-----LQVMDKIYVAGQIGLRI	111
Db	89	ELMINDTLAENVTYKQSFERYGLDHEIQGIYTHNPYIWNTPINVFQAVVYSLGRHDVMV	148	Db	89	ELMINDTLAENVTYKQSFERYGLDHEIQGIYTHNPYIWNTPINVFQAVVYSLGRHDVMV	148
QY	112	ILDRHP-----DCSGQ--SALWTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD	165	QY	112	ILDRHP-----DCSGQ--SALWTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD	165
Db	149	ILDNHLTKPGWCCANDGNGFFGQDFPTVVAALKQMAATFNGVSNVVGMSLRNELR-	207	Db	149	ILDNHLTKPGWCCANDGNGFFGQDFPTVVAALKQMAATFNGVSNVVGMSLRNELR-	207
QY	166	PACWGCDDPSIDWRLAARAGNAVLSVNPENLIFVEGVSYNGDSYWWGNGLQAGQYFV	225	QY	166	PACWGCDDPSIDWRLAARAGNAVLSVNPENLIFVEGVSYNGDSYWWGNGLQAGQYFV	225
Db	208	-----GYNHTSKDWKYMQKGAEAHVHTSNPNVILSGLN-----FDADLSFLKDRPV	255	Db	208	-----GYNHTSKDWKYMQKGAEAHVHTSNPNVILSGLN-----FDADLSFLKDRPV	255
QY	226	VLVNPNRLVYSADHYATSVGFQWFSDFPTFPNNMPG-IWNK---NWGYLPNQNIAPVWL	283	QY	226	VLVNPNRLVYSADHYATSVGFQWFSDFPTFPNNMPG-IWNK---NWGYLPNQNIAPVWL	283
Db	256	NLSPFKKLVLELHWYSFDTGTGQWKSHNVNDFCSQMFSEKERRTCGFLVDQGF-PLFLSE	314	Db	256	NLSPFKKLVLELHWYSFDTGTGQWKSHNVNDFCSQMFSEKERRTCGFLVDQGF-PLFLSE	314
QY	284	GT-----TLQSTTDTQTLVQLRPTAQXGAD-SFQWTFWS-----WNPDSGDTG----	329	QY	284	GT-----TLQSTTDTQTLVQLRPTAQXGAD-SFQWTFWS-----WNPDSGDTG----	329
Db	315	GTQQRGGDLEGN-----RYMNCMLAAEAKDLDAWAVAVTGVYFYREGKRGVVEA	364	Db	315	GTQQRGGDLEGN-----RYMNCMLAAEAKDLDAWAVAVTGVYFYREGKRGVVEA	364

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H. as-Neto, E.; Docena, C.; El-Dorray, H.; Facinani, A.P.; Ferreira, A.J.S. submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigz chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E. A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A. Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.F.; de Sa, R.G.; Santelli, R.V.; Sawasak M.; Tshahoko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z A:Reference number: A59328  
 C:Contents: annotation  
 C:Genetics:  
 A:Gene: XF0810

Query Match 16.5%; Score 328; DB 2; Length 565;  
 Best Local Similarity 29.6%; Pred. No. 7.6e-18;  
 Matches 103; Conservative 54; Mismatches 135; Indels 56; Gaps 12;

QY 12 EILDANNV-----PVRIAGNWFGETCNVYVHGLMSRDYRGMLOQISLGYNTIRL 63  
 Db 4 ETMDAQNVSDFHVEPLH--GVNWEGLTAQHLQLQDQRPVREVLDMQSLGINAIRL 61  
 QY 64 PYSDDILKPTGTPNSINFYQMNQDLOGLTSLQVMDKIVAYACIGIRIILDRHDPDC-SG 122  
 Db 62 PLCSDTLH-GTMTPTNLDLVR-NPDLKGRALTQIADAIIDEAGKGRVLLAYHGVCEPTD 119  
 QY 123 QSALWTSSVSATWISDLOALAQRYKGNPTTV-GFDLHNEPHDPACGCGPESIDWRLA 181  
 Db 120 GNPLLRSVDESEHQLSDVQFTTSHYRAQKQVGMVGLADMANHRRPFQSGDSTPDNRV 179  
 QY 182 ABRAGNAVLSVNPMLLIFVEGV-----QSYNGPSYWMGNLQAGQYVVLNPNR-LVYS 236  
 Db 180 VERAAAILANTPEWLGIVQVGLNPPCLDASIPSDNIQ--SCHCVPLRIPARNLLIM 237  
 QY 237 AHDYATSGVQPTWFSDFPTFNNMFGTWKNGWGLFNQNIAPVWLGEFG---TTLQSTTDQ 293  
 Db 238 PHFAGTDIDTE-----AALGAFSGKQTVLPSLDA 267  
 QY 294 TWLKTIVQLVRLPTAQYAGDSFQWTFMSNPNDSGDTGGILKDDWQTVDT 341  
 Db 268 TDAEQLAHLRIDALLAGIR--QGFYGSWTSQAQMPFGMLDNDGRTPRT 313

RESULT 13  
 S19652  
 C:Species: Pseudomonas fluorescens  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 15-Oct-1999  
 C:Accession: S19652; S16849  
 R:Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.  
 Biochem. J. 279, 793-799, 1991  
 A:Title: The cellobiohydrolase from *Pseudomonas fluorescens* subsp. *cellulosa* consists of p A:Reference number: S19652; MUID:92061996; PMID:1953673  
 A:Accession: S19652  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-748 <FER>  
 R:Ferreira, L.M.A.; Hazlewood, G.P.; Barker, P.J.; Gilbert, H.J.  
 Submitted to the EMBL Data Library, August 1991  
 A:Description: The cellobiohydrolase from *Pseudomonas fluorescens* subsp. *cellulosa* consist A:Reference number: S16849  
 A:Accession: S16849  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-262, 'C', 264-291, 'K', 293-748 <PE2>  
 A:Cross-references: EMBL:X61299; NID:945501; PID:945502  
 C:Genetics:  
 A:Gene: celC  
 C:Superfamily: bacterial cellulose-binding domain homology; glycosidase GWGW domain homo F;38-134/Domain: bacterial cellulose-binding domain homology <BCB>

F;183-220/Domain: glycosidase GWGW domain homology <GWG>  
 F;39-133/Disulfide bonds: #status predicted

Query Match 14.7%; Score 291.5; DB 2; Length 748;  
 Best Local Similarity 23.4%; Pred. No. 7.9e-15;  
 Matches 111; Conservative 61; Mismatches 138; Indels 165; Gaps 23;

QY 7 HTSGREILDANNVPVRIAGINWFGFTCTN-----YVHGLW-----S 43  
 Db 290 NTQGNLTQDQLPARG--NWFGLGRHEPSNDADNPSGAPMELYAGNNWVNNVNSQSGS 347  
 QY 44 RDRSRMLDQIKSYNTIRLPIYSDDIL-----KPTWPNISINFYQMNQDLOGLTSLQVMDK 99  
 Db 348 RTIQQTMTLTKQGGITMLRLPIAPQTLDANDPQGRSPNLKN-----HCSIRQSNARQALD 403  
 QY 100 IVAYAGQIGLRILDRH-----RPDCSGQSALWY-----TSSVS 133  
 Db 404 FIKLADQNDIQIFDIHSCSNYVWGRAGRLDARPPYDANRVGYDTFREYSCSANTNPS 463  
 QY 134 EAT-----WISDLOALA--QRYKGNPTVGVFDLHNEPHDPACGCGDPSIDWRLAA 182  
 Db 464 SVTRIHAYDKQKMLANLREIAGLSAKLGVSNLIGIDVFNPEYD-YTWA-----EWKGMV 516  
 QY 183 ERAGNAVLSVNPMLLIFVEGVQSYNGDSY-----WNGGNLQAGQY 223  
 Db 517 EAYQAINENVPNNMLLIVEGI--SANATQDGTPTSVVPVPHGSTDLNPNWGENLYEAGAN 575  
 QY 224 PVVLNVP-NRLVSAHDYATSVGPQTWFSDFPT-----FPNNMPG 261  
 Db 576 PP--NIFKDRLLFSPHTYGPVVFQRFQMDPAQTECAGLEGDEAAQACRQV-NPTVLEQ 633  
 QY 262 IWNKNGWGLFNQNIAPVWLGEFG-----TTLQSTTDQTLKTLVQYL 303  
 Db 634 GWSEHFGYLRGLGVG--ILIGFEGNMDWPKAKSQADRNWSHITTNVDQWQQAASYSF 692  
 QY 304 RPTAQYAGDSFQWTFMSNPNDSGDTGGILKDDWQTVDTVKDYLAPIKSSIFDPV 358  
 Db 693 K---RKGINA--CYWSMNPESADTMG-----W-----YLTP-----WDPV 722

RESULT 14  
 E86265  
 C:Species: Arabidopsis thaliana  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: E86265  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I. ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: E86265  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-522 <STO>  
 A:Cross-references: GB:A8005172; NID:94850396; PIDN:AAD31066.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F3F19.15  
 A:Map position: 1

Query Match 13.0%; Score 258.5; DB 2; Length 522;  
 Best Local Similarity 25.2%; Pred. No. 1.8e-12;  
 Matches 93; Conservative 62; Mismatches 153; Indels 61; Gaps 18;

QY 8 TSGREILDANNVPVRIAGINWFGFTCTNYYVHGLMSRDYRGMLOQISLGYNTIRLPSVD 67  
 Db 7 TSSRWIVDENGRLVKLVCAW--PSHLPQVWAEGLSKQPDVAVAKKIVEMGFCVRLTWPL 65

Db 343 YAOFWFKDFTMOTLLDDYWDYWAYTHDQIAPILIGWGGHMDGKQKQWMLLRDYI 402  
 Qy 304 RPTAOYCADSFQWTFWNSPDGTCGLKDDKQTVDTVKDGYLAP 349  
 Db 403 -----VONRIHTFWCINPSNGDTGLLGNLWSTWDEAKYALLKP 442

## RESULT 10

B47093  
 Cellulase (EC 3.2.1.4) CenD - Cellulomonas fimi  
 N:Alternate names: endo-1,4-beta-glucanase  
 C:Species: Cellulomonas fimi  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 06-Dec-1996  
 R:Accession: B47093  
 R:Mainke, A.; Gilkes, N.R.; Kilburn, D.G.; Miller Jr., R.C.; Warren, R.A.  
 J. Bacteriol. 175, 1910-1918, 1993  
 A:Title: Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase D (CenD),  
 A:Reference number: A47093; MUID:93209933; PMID:8458833  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-747 <MEI>  
 A:Experimental source: ATCC 484  
 A:Note: sequence inconsistent with the nucleotide translation  
 C:Function:  
 A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
 A:Pathway: cellulose degradation  
 C:Superfamily: bacterial cellulose-binding domain homology  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F:644-745/Domain: bacterial cellulose-binding domain homology <BCB>  
 F:645-744/Disulfide bonds: #status predicted

Query Match 27.7%; Score 549.5; DB 2; Length 747;  
 Best Local Similarity 35.0%; Pred. No. 5.3e-35;  
 Matches 132; Conservative 58; Mismatches 142; Indels 45; Gaps 15;  
 Qy 1 AGGGYW-HTSGREILDANNVPVRIAGINWGFETCNVYVHGLMSRDYRSMLOIKSLGN 59  
 Db 40 ATGDDMLHVEGNTIVDTSGEAILSLGVNWFNFGFNASERVFHGLWSGNITQTQQMAQRGIN 99  
 Qy 60 TIRLPYSDDL---KPGTW--PNSINFYQMNQDLQGLTSQVMDKIVAVAGQIGLRIILD 114  
 Db 100 VVRVPVSTQLLEKAGKFTLKP-NVNTY-ANPELEGKNSIQIPEYMLTLQCKYGIKVFLD 157  
 Qy 115 RH--RPPCSQ-SALWTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPH-----D 165  
 Db 158 VHSAEADNSGRVYNNWKKGDTTDDYVEGEWAATRWKDDDTIVGADIKNEPHGTQGST 217  
 Qy 166 PACWCGGDPSPIDWLAERAGNAVLSVNPMLLIFVEGVQSYN-----GDSY--W 212  
 Db 218 RAKWDGTTDKNFKHFAETASKILANPNWLVEVEGEVYPRPGVPWSTGTDYGTW 277  
 Qy 213 WGNLQAGQYPPVVLN-PPNRLYSADHYATSGPQTFP-SDTFPPNMPGINWKNWGYL 270  
 Db 278 WGNLGRVDRHPIDLGAHQDLQVYSPHDYGLVFDQKWFQKDFDKASLTADVWGNPNWLF 337  
 Qy 271 FNQNIAPWVLGERGTTI--QSTTQDW---LKTIVQYLRPTAQYCADSFQWTFWNSPDG 326  
 Db 338 HDEIAPLLIGEGGRGLQDPDRQDKMWAALRDVAER-----LSQTFVVLNPSG 388  
 Qy 327 DTGGLKDDMQTVDTVK 343  
 Db 389 DTGGLLLDDKWTDEVK 405

## RESULT 11

CZCUBM  
 Cellulase (EC 3.2.1.4) B precursor - Clostridium thermocellum  
 N:Alternate names: endo-1,4-beta-glucanase B precursor  
 C:Species: Clostridium thermocellum  
 C:Date: 31-Dec-1989 #sequence\_revision 31-Dec-1989 #text\_change 18-Jun-1999

C:Accession: A23512  
 R:Grepinet, O.; Beguin, P.  
 Nucleic Acids Res. 14, 1791-1799, 1986  
 A:Title: Sequence of the cellulase gene of Clostridium thermocellum coding for endoglucan  
 A:Reference number: A23512; MUID:86148508; PMID:3453102  
 A:Accession: A23512  
 A:Molecule type: DNA  
 A:Residues: 1-563 <GRE>  
 A:Cross-references: GB:X03592; NID:940668; PIDN:CAA27266.1; PID:940669  
 C:Comment: This secretory enzyme is part of a highly active and thermostable cellulase co  
 C:Genetics:  
 A:Gene: celB  
 C:Function:  
 A:Description: catalyzes the hydrolysis of 1,4-beta-D-glucosidic bonds in beta-D-glucans  
 A:Pathway: cellulose degradation  
 C:Superfamily: cellulase B; Clostridium cellulase repeat homology  
 C:Keywords: duplication; extracellular protein; glycosidase; hydrolase; polysaccharide de  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:26-563/Product: cellulase B #status predicted <MAT>  
 F:502-525/Domain: Clostridium cellulase repeat homology <CCR1>  
 F:534-557/Domain: Clostridium cellulase repeat homology <CCR2>  
 Query Match 26.3%; Score 522; DB 1; Length 563;  
 Best Local Similarity 30.7%; Pred. No. 5.1e-33;  
 Matches 129; Conservative 58; Mismatches 159; Indels 74; Gaps 13;  
 Qy 1 AGGGY-----W-HTSGREILDANNVPVRIAGINWGFETCNVYVHGLMSRDYRSM 50  
 Db 27 AEGSYADLAEPDDDDMLHVEGNTIVDKYGNKYMITGANWFGNCRERMLDLSYHSDIADI 86  
 Qy 51 DOIKSLGYNVIRLPYSDDL---KPGTWPNNSINFYQMNQDLQGLTSQVMDKIVAVAGQI 107  
 Db 87 ELVADKGINVVRMPFIATDLLYANSGQIYPPSTDITSYNNPALAGLSYELFNMLENFKRV 146  
 Qy 108 GLRIILDRHDPDCSQS---ALWTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPH 164  
 Db 147 GIKVILDVHSPETDNOGHNYPLWYNTTITEIFKAWVWAERYKNDTITGIFDLKNEPH 206  
 Qy 165 -----DPACWCGGDPSPIDWLAERAGNAVLSVNPMLLIFVEGVQSY----- 206  
 Db 207 TWGTMKIKASAIWDDSNHNPWKRVABETALAILEVHPNVLIFVEGVEMYPKDGIIWD 266  
 Qy 207 -----NGDSY--WGNLQAGQYPPVL-NVPNRLYSADHYATSGPQTFWSDP 253  
 Db 267 EFTFDTSPWTGNDYGNWCGNLRGVDPINLGYSQVSPHDYSPHDYPIVEQDFKGD 326  
 Qy 254 FPPNN-----MPGINWKNWGYLFPNQIAPVLMGERGTTIQTQDTWLTQVYLRP 305  
 Db 327 FITANDEQAKILYEQCWRDNWAYIMEGISTPLLEGWGMTEG--GHPLDLMLKYLRC 384  
 Qy 306 TAAQYCADS---FQWTFWNSPDGTCGLKDD-----WQTVDTVKDGYLAPI 350  
 Db 385 MRDFTLENKYKLHHTFWCINIDSADTGCLFTRDEGTPPGGRDLKWN--DNKYDNYLYPV 442

## RESULT 12

B82761  
 extracellular endoglucanase precursor XF0810 [imported] - Xylella fastidiosa (strain 9a5c  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 A:Accession: B82761  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequenc  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82761  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-565 <SIM>  
 A:Cross-references: GB:AE003920; GB:AE003949; NID:99105697; PIDN:AA783620.1; GSPDB:GN0012  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al

F:20-320/Domain: Streptomyces endo-1,4-beta-xylanase A homology <SXY>

Query Match	31.4%;	Score 624.5;	DB 2;	Length 915;
Best Local Similarity	35.5%;	Pred. No. 9.6e-41;		
Matches 133;	Conservative 68;	Mismatches 135;	Indels 39;	Gaps 12;
QY	6	W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDRYSMLDQIKSLGYNTRILP	64	
Db	505	WLYVSGNKIVDXDGRFWLTGYNWFGYNTGTNVFDGWSNCLKDILAEIANRGENLLRVP	564	
QY	65	YSDIL---KPGTMEN-SINFGQMNDLOGLTSLQVMDKIVAYAGQIGLRIILDSH--RP	118	
Db	565	ISAEILNWSQGIYKPKPNINY-VNPELEGKNSLEVFIVQTCKEVGLKIMLDIHSIKT	623	
QY	119	DCSGQ-SALWYTSVSEATWISDLOALAQRYKGNFTVVGFDLHNEPHDP-----ACWG	170	
Db	624	DAMGHYFVWYDEKFTPEDFYKACETWIRYKNDTIIAFDLKNEPHGKFWODTTFAKWD	683	
QY	171	CGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWGGNL	217	
Db	684	NSTDINNWKYAAETCAKRIININPNLLIVIEGIEAYPKDDVTWTSKSSDYSTWGGNL	743	
QY	218	QGAGQYPVVL-NVFNRLVYSAHDYATSGPOTWFSDFPNN--MPGIWKNKNGYLFNQ	274	
Db	744	RGVRKYPIINLGKYNQKVVSPHDYGSVYQQPWFY-PGFTKESLQDCWRPNWAYIMEEN	802	
QY	275	IAPVWLGEFGTTLQSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGLKD	334	
Db	803	IAPLLIGEWGCHLDGADNEKWKYLRDYII-----ENHIHTFWCFNANSODTGLGVY	856	
QY	335	DWQTVDTVKGYLAP 349		
Db	857	DFTTWDEKYSFLKP 871		
RESULT 8				
S02711	cellulase (EC 3.2.1.4) precursor - Caldocellum saccharolyticum			
N:Alternate names:	endo-1,4-beta-glucanase			
N:Contains:	cellulase (EC 3.2.1.4); cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91)			
C:Species:	Caldocellum saccharolyticum			
C>Date:	07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999			
C:Accession:	S02711			
R:Saul, D.J.; Williams, L.C.; Love, D.R.; Chamley, L.W.; Bergquist, P.L.				
Nucleic Acids Res.	17, 439, 1989			
A:Title:	Nucleotide sequence of a gene from Caldocellum saccharolyticum encoding for exo			
A:Reference number:	S02711; MUID:89098398; PMID:2789517			
A:Accession:	S02711			
A:Molecule type:	DNA			
A:Residues:	1-1039 <SAU>			
A:Cross-references:	EMBL:X13602; NID:g40645; PIDN:CAA31936.1; PID:g40646			
C:Genetics:				
A:Gene:	celB			
C:Function:				
A:Description:	hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce			
A:Pathway:	cellulose degradation			
C:Superfamily:	Streptomyces endo-1,4-beta-xylanase A homology			
C:Keywords:	glycosidase; hydrolase; polysaccharide degradation			
F:1-28/Domain:	signal sequence #status predicted <SIG>			
F:29-1039/Product:	cellulase #status predicted <NAT>			
F:72-373/Domain:	Streptomyces endo-1,4-beta-xylanase A homology <SXY>			
Query Match	31.4%;	Score 624.5;	DB 2;	Length 1039;
Best Local Similarity	35.5%;	Pred. No. 1.1e-40;		
Matches 133;	Conservative 68;	Mismatches 135;	Indels 39;	Gaps 12;
QY	6	W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDRYSMLDQIKSLGYNTRILP	64	
Db	629	WLYVSGNKIVDXDGRFWLTGYNWFGYNTGTNVFDGWSNCLKDILAEIANRGENLLRVP	688	
QY	65	YSDIL---KPGTMEN-SINFGQMNDLOGLTSLQVMDKIVAYAGQIGLRIILDSH--RP	118	
Db	689	ISAEILNWSQGIYKPKPNINY-VNPELEGKNSLEVFIVQTCKEVGLKIMLDIHSIKT	747	

QY	119	DCSGQ-SALWYTSVSEATWISDLOALAQRYKGNFTVVGFDLHNEPHDP-----ACWG	170	
Db	748	DAMGHYFVWYDEKFTPEDFYKACETWIRYKNDTIIAFDLKNEPHGKFWODTTFAKWD	807	
QY	171	CGDPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSYNGD-----SYWGGNL	217	
Db	808	NSTDINNWKYAAETCAKRIININPNLLIVIEGIEAYPKDDVTWTSKSSDYSTWGGNL	867	
QY	218	QGAGQYPVVL-NVFNRLVYSAHDYATSGPOTWFSDFPNN--MPGIWKNKNGYLFNQ	274	
Db	868	RGVRKYPIINLGKYNQKVVSPHDYGSVYQQPWFY-PGFTKESLQDCWRPNWAYIMEEN	926	
QY	275	IAPVWLGEFGTTLQSTTDQTLKTLVQYLRPTAQYGADSFQWTFWSNPNDSGDTGGLKD	334	
Db	927	IAPLLIGEWGCHLDGADNEKWKYLRDYII-----ENHIHTFWCFNANSODTGLGVY	980	
QY	335	DWQTVDTVKGYLAP 349		
Db	981	DFTTWDEKYSFLKP 995		
RESULT 9				
A40589	cellulase (EC 3.2.1.4) - Clostridium thermocellum			
N:Alternate names:	endo-1,4-beta-glucanase			
C:Species:	Clostridium thermocellum			
C>Date:	07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jun-2000			
C:Accession:	A40589; S31381			
R:Lemaire, M.; Beguin, P.				
J. Bacteriol.	175, 3353-3360, 1993			
A:Title:	Nucleotide sequence of the celG gene of Clostridium thermocellum and characteri			
A:Reference number:	A40589; MUID:93273701; PMID:8501039			
A:Accession:	A40589			
A:Status:	preliminary			
A:Molecule type:	DNA			
A:Residues:	1-566 <LEM>			
A:Cross-references:	GB:X69390; NID:g40677; PIDN:CAA49187.1; PID:g40678			
C:Genetics:				
A:Gene:	celG			
C:Function:				
A:Description:	hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel			
A:Pathway:	cellulose degradation			
C:Superfamily:	cellulase B; Clostridium cellulase repeat homology			
C:Keywords:	glycosidase; hydrolase; polysaccharide degradation			
F:503-526/Domain:	Clostridium cellulase repeat homology <CCR>			
F:536-559/Domain:	Clostridium cellulase repeat homology <CCR>			
Query Match	30.2%;	Score 600.5;	DB 2;	Length 566;
Best Local Similarity	34.2%;	Pred. No. 3.8e-39;		
Matches 139;	Conservative 55;	Mismatches 143;	Indels 69;	Gaps 15;
QY	6	W-HTSGREILDANNVPVRIAGINWFGFETCNVYVHGLMSRDRYSMLDQIKSLGYNTRILP	64	
Db	44	WLHCKGNKIYDMYGNEVLTGANWFGFNCSECFHGAW-YDVKTILTSIADRGINLLRP	102	
QY	65	YSDIL-----KPG-----TMENSINFQMNODLQ-----LTSLOWMDKIVAYAGQI	107	
Db	103	ISTELLYSNWIGKPNVSSVTASNPPYHVVPDFDPETDDVKNSEWIFDIIMGCKEL	162	
QY	108	GLRIILDRHPDC--SQQS-ALWY---TSS---VSEATWISDLOALAQRYKGNFTVVGFD	158	
Db	163	GIXVMIDIHSPDANNSCHNYELWYKSTSTCGVVTWKWIDTLVWLADKYENDDTIAFD	222	
QY	159	LHNEPHDP-----ACWCGGSPSIDWRLAERAGNAVLNVNPNLLIFVEGVQSY- 206		
Db	223	LKNEPHGKGYTAEVFKLLAKWNSDNNKYAAETCAKRIILNPNLLIVIEGVQSY- 282		
QY	207	-----NGD-----SYWGGNLQAGQYFVVLNPN-NLVYSADHYATS- 244		
Db	283	KTEKGYTDTFDIWGATGDASPWYSAAWGNLGRGVKDYDIDLGPLNSQIVYSPHDYGPSV	342	
QY	245	GPQTWP-SDTFFNNPFGIWNKNGYLFNQIAPVWLGEFGTTLQSTTDQTLKTLVQYL	303	

QY 220 AGOYFVVL-NVFNRLVYSAHYATSVGQTFW-SDPTFPNNPMGIWNKNGYLFNQNIAP 277  
 DB 263 VKDYFVDLAPYKNOVYSPHDYGPVSDQTFWDCGDFTEQSLLNDIKRPSFYIQEKNIAP 322  
 QY 278 VWLGFEGTTLQSTDTQWTLKTLVQLVRLPTAQYAGDSFQWTFWSNPSGDTGILKDDWQ 337  
 DB 323 LLIIGWGGNMGCKNEQWMTMAKLI-----SDKNMHTFWCLNANSQDTGILLEYDFX 376  
 QY 338 TVDTVTKDGYLAP 349  
 DB 377 TIDTKKALVOP 388  
 RESULT 5  
 E82759  
 endo-1,4-beta-glucanase XF0818 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: E82759  
 R:anonymus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; PMID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: E82759  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-592 <SIM>  
 A:Cross-references: GB:A8003921; GB:A8003949; NID:g9105710; PIDN:AAF83628.1; GSPDB:GN001  
 R:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
 as-Neto, E.; Docena, C.; M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krueger, J.E.; Kuran, E.B.; Laigh  
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Mencia, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF0818  
 Query Match 34.6%; Score 687.5; DB 2; Length 592;  
 Best Local Similarity 42.6%; Pred. No. 6.3e-46;  
 Matches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;  
 QY 5 YWHTSREILDANNVRIAGINWFGTETCNVYVHGLWSRDYRSMLDQIKSLGTYNIRLP 64  
 DB 26 YSIHGKVIDDKGN-QIOLKGSWFGFTTNVYVHGLWTRNWKKEFITQIOSMGLNAVRLLP 84  
 QY 65 YSDDLKPGTFNPNINFYQNNQDLQGLTSLQVMDKIVAVAGQIGLRIIDRRHPCGSGS 124  
 DB 85 FCPASLNSNTSPSSID-YNRPDLQLGLSLLIQIMDKVKVLSDRGIYVLDHHTPDCAATIS 143  
 QY 125 ALWTSSTVSEATWISDQALAAQRYKGNPTVVGFDLNEPHDPACWCGGSDPISDWLAAR 184  
 DB 144 ELWYDTSYSEKQWIDDLRFVAHRYANVPGVIGLDVKNPHGRATWGTGPKTDWNTAVEH 203  
 QY 185 AGNAVLSVNPMLLIFVEGVSQSYNGDS-----YWGNGNLQAGQYFVVLNVP-NRLVYSAHD 239  
 DB 204 AAAAILEAAPKWLIVGEGICENPSCSSTIGHFWGENLEPMDCFP-LKVPADHLLILPHV 261  
 QY 240 YATSVGQTFWSDPTFPNNPMGIWNKNGYLFNQNIAPVWLGFGTTL--QSTTDTWLK 297  
 DB 262 YGPDVYVQYFNSPDFPNNMAAIDWRFHFAKAGYA-MAIGFSGKYGKGEGRDVAQON 320  
 QY 298 TLVQLRPTAQY-ADSFQWTFWSNPSGDTGILKDDWQTV 339

DB 321 ALVDYL---ISIGVTDTF---YWSINANSSTGGLFRDDMNHV 357

## RESULT 6

E75142  
 endoglucanase PAB0632 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000

C:Accession: E75142

R:anonymus, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc

A:Reference number: A75001

A:Accession: E75142

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-514 <KAW>

A:Cross-references: GB:AJ248285; GB:AL096836; NID:g5458067; PIDN:CAB49854.1; PID:g545836

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: celB-like; PAB0632

C:Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 34.3%; Score 682.5; DB 2; Length 514;

Best Local Similarity 41.6%; Pred. No. 1.3e-45;

Matches 148; Conservative 57; Mismatches 122; Indels 29; Gaps 10;

QY 7 HTSREILDANNV-----PVRTAGINWFGTETCNVYVHGLWSRDYRSMLDQIKSLGTYN 60

DB 30 YTAENGIFVQNVITGKGLYDHGVSFGFELKDHVYGLDKRNWKDILKDVGLGFNA 89

QY 61 IRLPYSDILKPGTTPN--SINFYQNNQDLQGLTSLQVMDKIVAVAGQIGLRIIDRRHP 118

DB 90 IRLPFCSIESIRPTDRSPERIN-YELNPDKNLTSLEIMEKIEVANSIGLYLLDYHRI 148

QY 119 DCSSQALWTSSTVSEATWISDQALAAQRYKGNPTVVGFDLNEPHDPACWCGGSDPISDW 178

DB 149 GCEIEFLWTENYSEEQYKDWIFLAKRFGKYPNVIGADIKNEPHGEAGWGTGDER-DF 207

QY 175 RLAAERAGNAVLSVNPMLLIFVEGVSQSYNGDSYV--WGGNLQAGQYFVVL 227

DB 208 RLFAEKVGRILKVAHPHLIFVETQYTHVPNIDEIEKKGWTFWGENLMGVKDYFVRL 267

QY 228 NVNRLVYSAHYATSVGQTFWSDPTFPNNPMGIWNKNGYLFNQNIAPVWLGFGTTL 287

DB 268 -PRGKVYSPHYVGPVYVMDYFSPDPNNPILWETHFGYLTDLNLTLYV-IGWGGNY 325

QY 288 QSTTDTQWTLVQLRPTAQYAGDSFQWTFWSNPSGDTGILKDDWQTVDTVK 343

DB 326 EG-LDKVQWQAFVKWLKKIY-----NFFYCLNPESGDTGILFDDWKTYNWEK 375

## RESULT 7

A43802

cellulase (EC 3.2.1.4) / cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - Caldocellum se

N:Alternate names: endo-1,4-beta-glucanase

C:Species: Caldocellum saccharolyticum

C:Date: 30-Jan-1993 #sequence\_revision 30-Sep-1993 #text\_change 10-Jul-1998

C:Accession: A43802

R:Saul, D.O.; Williams, L.C.; Grayling, R.A.; Chamley, L.W.; Love, D.R.; Bergquist, P.L.

Appl. Environ. Microbiol. 56, 3117-3124, 1990

A:Title: celB, a gene coding for a bifunctional cellulase from the extreme thermophile "C

A:Reference number: A43802; PMID:91136262; PMID:2126700

A:Accession: A43802

A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual trar

A:Molecule type: DNA

A:Residues: 1-915 <SAU>

A:Cross-references: EMBL:X13602

C:Function:

A:Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel

A:Pathway: cellulose degradation

C:Superfamily: Streptomyces endo-1,4-beta-xylanase A homology

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

JH0158  
 cellulase (EC 3.2.1.4) precursor - *Xanthomonas campestris* pv. *campestris*  
 N/Alternate names: endo-1,4-beta-glucanase; extracellular endoglucanase  
 C/Species: *Xanthomonas campestris* pv. *campestris*  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 24-Sep-1999  
 C/Accession: JH0158  
 R/Gough, C.L.; Dow, J.M.; Keen, J.; Henrissat, B.; Daniels, M.J.  
 Gene 89, 53-59, 1990  
 A/Title: Nucleotide sequence of the *engXCA* gene encoding the major endoglucanase of *Xanthomonas campestris* pv. *campestris*  
 A/Reference number: JH0158; MUID:90323605; PMID:2373365  
 A/Accession: JH0158  
 A/Molecule type: DNA  
 A/Residues: 1-493 <GOU>  
 A/Cross-references: GB:X32700; NID:g155397; PIDN:AAA27612.1; PID:g155398  
 C/Comment: Endoglucanase plays a minor role in the early stages of pathogenicity of *Xanthomonas campestris* pv. *campestris*  
 C/Genetics:  
 A/Genes: *engXCA*  
 C/Function:  
 A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cellulose  
 A/Pathway: cellulose degradation  
 C/Superfamily: *Xanthomonas campestris* cellulase  
 C/Keywords: glycosidase; hydrolase; polysaccharide degradation  
 F/1-25/Domain: signal sequence #status predicted <SIG>  
 F/26-474/Product: cellulase #status predicted <MAT>  
 F/375-399/Region: proline/threonine-rich

Query Match	40.3%	Score	800.5;	DB	1;	Length	493;
Best Local Similarity	47.2%	Pred.	No. 7.2e-55;				
Matches	162;	Conservative	58;	Mismatches	104;	Indels	19; Gaps
							9;

QY 5 YHSTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNIRLP 64  
 DB 26 YSINNSRQIVDDSGKVQLKGVNFGFETGNVHVMGLWAARWKMIMVQVQGLGFNAVRLP 85  
 QY 65 YSDILKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLIILDRHPDCSGQS 124  
 DB 86 FCPATLSDTNPASID-YSRNADLQGLTSLQILDKVIAEFNARGVMYVLLDHPHTPCAGIS 144  
 QY 125 ALWYTSVSSEATWISDQALQAKYKGNPTVVGFDLHNEPHDPACWCGCDPSIDWRLAAER 184  
 DB 145 ELWYTGSTYEAQLADLRFVANRYKVPVVLGLDKNEPHGAATWGTGNATDNKAAER 204  
 QY 185 AGNAVLSVNNLLIFVEGVO-----SYNGDSYWWGNLQAGQYPPVVLNVP-NRLVYSAH 238  
 DB 205 GSAVLAVAPKWLIAVSGITDNPVCSTNG-GIEWGNLQPLACTP-LNIPANRLLIAPH 261  
 QY 239 DYATSVGPOTWFSDEPTENMEGWNKMGVLFNQNTAPVWLGEFGTTL--QSTTDQOTWL 296  
 DB 262 VYGEDVVFQIFDNSNPNMIPAWEHFQFAGTH--ALLGEFGKYGEGDARDKTWQ 319  
 QY 297 KTLVQYLRPTAQYGADSFQNTFWSNPDSGTGGILKDWQTV 339  
 DB 320 DALVKYLR---SKGIN--QGFWGNPNFSGDTGGILRDDWTSV 357

RESULT 3  
 E71059  
 probable endo-1,4-beta-glucanase - *Pyrococcus horikoshii*  
 C/Species: *Pyrococcus horikoshii*  
 C/Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
 C/Accession: E71059  
 R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekinaka, M.; Ohtsuka, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Oguchi, N.  
 M. Res. 5, 55-76, 1998  
 A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon  
 A/Reference number: A71000; MUID:98344137; PMID:9679194  
 A/Accession: E71059  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-458 <XAW>  
 A/Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30271.1; PID:g3257588  
 A/Experimental source: strain OT3  
 A/Note: this accession replaces an interim accession for a sequence replaced by GenBank

C;Genetics:  
A;Gene: Phil71  
C;Superfamily: cellulase B; Clostridium cellulase repeat homology

Query Match 38.0%; Score 755; DB 2; Length 458;  
Best Local Similarity 45.0%; Pred. No. 2.4e-51;  
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVPVRIAGINWFFTCNNVHGLWSRDYRSMLOQIKSLGYNTIRLPYS 67  
DB 54 TSCEE-----TPIHLFGVNWFGFETPNVHGLWKRNMEDMLLQIKSLGNAIRLPCT 107  
QY 68 DILKPGCTMPSINFYOMNODLGLTSLQVMDKIVAYAGQIGLRIILDBHRPDCSQSALW 127  
DB 108 ESKVPTQPIGID-YKMPDURLGLDSLOIMEKIKKAGDLGIFVLLDYHRTGCTHIEPLW 166  
QY 128 YTSVSEATWISDLOALACRYKGNPTVVGFDLHNEPHD-----PACWGCDPS 175  
DB 167 YTEDSEEDFINTWIEAKRFKGYWNVIGADLKNEPHSVTSPPAAAYTDGTGATWGMG 226  
QY 176 IDWLLAABRAGNAVLSVNPENLLIFVEGVQSYN-----GDSYWGCGNLQAGQIPV 226  
DB 227 TQWNLAAERIGAKILKAPHLWIFVEGTQFTNPKTSSYKMGYNWAGGNLMAVKDYPV- 285  
QY 227 LNVP-NRLVYSADYATSVGPOTWFSDET-PPNNMPGTWKNKGYLEFNQNIAPVLMGEFG 284  
DB 286 -NLPRNKLVYSHVYCPDYINQYFGPAKGFNDLPDIWYHHFGYVKLELGSYVVGEEFG 344  
QY 285 TTL---QSTDTQTLKTLVQLYRLPFAQYAGDSFQWTFWSWNPDSGDTGGILKDKQTV 339  
DB 345 GRYGHGSDPRDIWONKLVDMW--IENKFCDF--YMSWNPDSGDTGGILQDDWTTI 397

RESULT 4  
E97012  
probable non-processive endoglucanase family 5, secreted, Cella homolog secreted, dockerr  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: E97012  
R;Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A;Reference number: A96900; PMID:21359325; PMID:21359325  
A;Accession: E97012  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-482 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK78888.1; PID:sl5023812; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC0912

Query Match 34.7%; Score 690; DB 2; Length 482;  
Best Local Similarity 37.9%; Pred. No. 3e-46;  
Matches 141; Conservative 71; Mismatches 126; Indels 34; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFFTCNNVHGLWSRDYRSMLOQIKSLGYNTIRLP 64  
DB 24 YLHSDGSKLLDDYGNQVRMTGIAWFELETNPYCFHGLWANRLDNLINIVADNGFNTLRVP 83  
QY 65 YSDTIL---KPGT--WPSINIFYQMNQDLOGLTSLQWMDKIVAYAGQIGLRIILDRHPD 119  
DB 84 LSVLEYNQWRQGYPTTPDSINDY-ISPCLKGONSILQLDDVIAYSKKVGVKVMGLMDHRIE 142  
QY 120 CSQGSALWTSVSEATWISDLOALACRYKGNPTVVGFDLHNEPHDPAC-----WGCG 172  
DB 143 SGGQATWTSKYTTDDYEKCQWYLADRYKNDTVIADIENEPHGKAYRAETSAKNDT 202  
QY 173 DSDIDWRLAABRAGNAVLSVNPENLLIFVEGVQSY-----NGDSY-----WVGGNLQ 219  
DB 203 TDEDNWRYEAEKVGKILLINPKMLIVVEGYETYPKEGTAAGSTNPDPDYYGWWGNNLRG 262

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: August 2, 2004, 16:30:59 ; Search time 12.3333 Seconds  
(without alignments)  
2792.154 Million cell updates/sec

Title: US-09-997-504A-10  
Perfect score: 1987  
Sequence: 1 AGGGYHTSGREILDANNVP.....VDTVKDGYLAPIKSSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1054	53.0	397	2 A35136	cellulase (EC 3.2.1.4) - Bacillus polymyxa
2	800.5	40.3	493	1 JH0158	cellulase (EC 3.2.1.4) - Bacillus polymyxa
3	755	38.0	458	2 E71059	probable endo-1,4-beta-glucanase
4	690	34.7	482	2 E97012	endo-1,4-beta-glucanase
5	687.5	34.6	592	2 E82759	endo-1,4-beta-glucanase
6	682.5	34.3	514	2 E75442	endo-1,4-beta-glucanase
7	624.5	31.4	915	2 A43802	cellulase (EC 3.2.1.4) - Bacillus polymyxa
8	624.5	31.4	1039	2 S02711	cellulase (EC 3.2.1.4) - Bacillus polymyxa
9	600.5	30.2	566	2 A40589	cellulase (EC 3.2.1.4) - Bacillus polymyxa
10	549.5	27.7	747	2 B47093	cellulase (EC 3.2.1.4) - Bacillus polymyxa
11	522	26.3	563	1 CZCLBM	extracellular endo-1,4-beta-glucanase
12	328	16.5	565	2 E82761	cellulase (EC 3.2.1.4) - Bacillus polymyxa
13	291.5	14.7	748	2 S19652	cellulase (EC 3.2.1.4) - Bacillus polymyxa
14	258.5	13.0	522	2 E86265	cellulase (EC 3.2.1.4) - Bacillus polymyxa
15	221	11.1	526	2 T51476	cellulase (EC 3.2.1.4) - Bacillus polymyxa
16	211.5	10.6	488	2 T51502	cellulase (EC 3.2.1.4) - Bacillus polymyxa
17	175	8.8	426	2 A42649	cellulase (EC 3.2.1.4) - Bacillus polymyxa
18	181.5	8.1	814	1 CZCLBM	extracellular endo-1,4-beta-glucanase
19	160.5	8.1	722	2 H96986	cellulase (EC 3.2.1.4) - Bacillus polymyxa
20	158.5	8.0	516	2 JE0134	cellulase (EC 3.2.1.4) - Bacillus polymyxa
21	155.5	7.8	356	2 G82523	cellulase (EC 3.2.1.4) - Bacillus polymyxa
22	148	7.4	1012	2 E97326	cellulase (EC 3.2.1.4) - Bacillus polymyxa
23	147.5	7.4	406	2 A43722	cellulase (EC 3.2.1.4) - Bacillus polymyxa
24	141.5	7.1	329	2 C72216	cellulase (EC 3.2.1.4) - Bacillus polymyxa
25	140.5	7.1	500	2 S22458	cellulase (EC 3.2.1.4) - Bacillus polymyxa
26	140.5	7.1	584	2 J01229	cellulase (EC 3.2.1.4) - Bacillus polymyxa
27	137.5	6.9	388	1 JC5461	cellulase (EC 3.2.1.4) - Bacillus polymyxa
28	137	6.9	370	2 G97001	cellulase (EC 3.2.1.4) - Bacillus polymyxa
29	137	6.9	900	2 JH0157	cellulase (EC 3.2.1.4) - Bacillus polymyxa

ALIGNMENTS

RESULT 1

A35136  
Cellulase (EC 3.2.1.4) - Bacillus polymyxa  
N:Alternate names: endo-1,4-beta-glucanase  
C:Species: Bacillus polymyxa  
C:Date: 03-Aug-1990 #sequence\_revision 03-Aug-1990 #text\_change 24-Sep-1998  
C:Accession: A35136  
R:Baird, S.D.; Johnson, D.A.; Seligy, V.L.  
J: Bacteriol. 172, 1576-1586, 1990  
A:Title: Molecular cloning, expression, and characterization of endo-beta-1,4-glucanase  
A:Reference number: A35136; MUID:90170877; PMID:2307659  
A:Accession: A35136  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-397 <BAI>  
A:Cross-references: GB:M33791; GB:M33840  
C:Function: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as cel  
A:Pathway: cellulose degradation  
C:Superfamily: Xanthomonas campestris cellulase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 53.0%; Score 1054; DB 2; Length 397;  
Best Local Similarity 54.5%; Pred. No. 8.1e-75;  
Matches 194; Conservative 44; Mismatches 106; Indels 12; Gaps 4;

QY	4	GYWHTSGREILDANNVPVRIAGINWFGPETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRL	63
DB	37	GYHTQGNKI VDESGKEAFAFNGLWFGLETPTNYTHGLWSRSMDDMLDQVKEGYNIRL	96
QY	64	PYSDDLPGTGPNSINFYQNQDLQGLTSLOVMDKIVAYAGQIGLRILDRHPCSGQ	123
DB	97	PYSNOLFSSGRPSID-YHNPDVLGNPIQIWDKLEKAGQGIQIILDRHPCSGQ	155
QY	124	SALWYTSVSSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAEE	183
DB	156	SELWYTSQYPSERWISDWKMLADRYKNNPTVIGADLHNEPHDQASWGTGASTDWRLAQ	215
QY	184	RAGNAVLSPNPLILFVGVQ---SYNGDSYMGWGNLOGAGQYPVNLNPNRLVVSADY	240
DB	216	RAGNAVLSPNPLILFVGVQ---SYNGDSYMGWGNLOGAGQYPVNLNPNRLVVSADY	275
QY	241	ATSVGPQWTFSDPTFPNNMGPINWKNWGLFNQNIAPVLMGFEFG--TTLSQTTDTQWTKT	298
DB	276	GFVGSQWQWENDPAPPSNLPALMDQWYISKONIAPIVLGVEFGGRNVDSLSPEGMQNA	335
QY	299	LVOYLRPTAQYGAQDSFQWTFWSNPDSDGTGILKDDQCTVDTVKDGYLAPIKSSI	354
DB	336	LVHYI-----GANNLYFTYSLNPNFSGDTGGLLDDDTTWNRPQDMLGRIMKPV	385

RESULT 2



This Page Blank (uspto)

```

; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9262
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9262

Query Match      33.0%; Score 655.5; DB 15; Length 421;
Best Local Similarity 39.9%; Pred. No. 3.1e-55;
Matches 141; Conservative 60; Mismatches 123; Indels 29; Gaps 12;

QY 5 YWHTSGREILDANNVPRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTRLRP 64
Db 26 YSISHGKVVDKGN-QIQLRGVNWFGFETGHHVVGVLWNRNKEFITQLQMGFNAILRP 84
QY 65 YSDDILKPGTWPNISINFYOMNODLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSGQS 124
Db 85 FCPANLNSNTSPSSID-YSRNPDLOGLSSLOILDKVVKELSDREMYVLLDHRPDCSAIS 143
QY 125 ALWYTSSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAER 184
Db 144 ELWHTDSYSEKOWIDDLRFVAHRYANVPGVIGLEVKNEPHGRATWGTGDFQTDWNTAVEH 203
QY 185 AGNAVLNVNPLLIFVSGVQSYNGDS-----YMWGSLQAGQYPPVVLNVP-NRLVY 235
Db 204 AAAAILEAPKXVIGIGI-----GENICSTIGHFNGENLEPMDCPT--LKVPADHLLL 257
QY 236 SAHDYATSVGPQTWFSPTFPNNMPGIWNKNWGYLFNQNTAPVWLGEFGTTL--QSTTDQ 293
Db 258 MPVYVGPVYVQYFNSGDFPNNMAAIWDKHPHFAGAGYA-MAIGFEGGKYGEGDPRDI 316
QY 294 TWLKITLYOYLRPTAQYG-ADSFQWTFWSNPNDSGDTGILKDDWQTV--DTVK 343
Db 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSGDTGGMVGNWMTTPRDDKVK 363

```

Search completed: August 2, 2004, 16:40:48  
Job time : 42.3333 secs

Best Local Similarity 42.6%; Pred. No. 3.7e-58;  
Matches 146; Conservative 50; Mismatches 128; Indels 19; Gaps 10;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLNSRDYRMLDQIKSLGNTIRLP 64  
Db 26 YSISHGKVIDDKGN-QIQKGGSWFGFETTNVHVLWTRNWKFTIQSGMLNAVRLP 84

QY 65 YSDIILKPGTMPNSINFYQMNODLOGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQS 124  
Db 85 FCPASLNSNTSPSID-YNRNPDQLQGLSLQIMDKVVKLSDRGIYVLMDHHTPDCAALS 143

QY 125 ALWYSSVSEATWISDLQALARYKGNPTVGFDLNHEPHDPACWCGDPSIDWELAAER 184  
Db 144 ELWYDYSYSEKQWIDDLRFVAHRYANVPVIGLVDKNEPHGRATWGTGDKTDMNTAVEH 203

QY 185 AGNAVLNVNPNLLIFVEGVQSYNGDS----YWGNGMLQAGQYPPVNLVNP-NRLVYSAHD 239  
Db 204 AAAAILEAAPKWLIGVEGIGENPSCSTIGHFNGENLEPMDCTP--LKVPAHLLIIEHV 261

QY 240 YATSGPQTFWSDPTFPNNMNGIWNQWGLYFNQNIAPVWLGEFGTTL--QSTTDQWLK 297  
Db 262 YGPDVYVQPYFNPSDPFPNMAAIDKHFGHFAKAGYA-MAIGFEGKYGEGDPRDVAQGN 320

QY 298 TLVQYLRLPTAQYG-ADSFQWTFWNNPDSGDTGGLKDDWQTV 339  
Db 321 ALVDYL---ISIGVTDTF---YWSINANSSTGGGLFRDDWNHV 357

## RESULT 13

US-10-369-493-21618  
; Sequence 21618, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 21618  
; LENGTH: 514  
; TYPE: PRT  
; ORGANISM: Pyrococcus abyssi  
US-10-369-493-21618

Query Match 34.3%; Score 682.5; DB 15; Length 514;  
Best Local Similarity 41.8%; Pred. No. 9.2e-58;  
Matches 148; Conservative 57; Mismatches 122; Indels 29; Gaps 10;

QY 7 HTSGREILDANNV-----PVRIAGINWFGFETCNVYVHGLNSRDYRMLDQIKSLGNT 60  
Db 30 YTAENGIIIFQNVTTCEKKPLYLHGVSWFGEFELKOHVYVGLDKRWKQILKDVKELGNA 89

QY 61 IRLPYSDDLKPGTMEN--SINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRIILDRHRP 118  
Db 90 IRLPFCSEIRPDTRSPERIN-YELNPDNLKNTLSLEIMEKIEYANSIGLYLLDYHRI 148

QY 119 DCSGQSALWYTSVSSEATWISDLQALARYKGNPTVGFDLNHEPHDPACWCGDPSIDW 178  
Db 149 GCEIIEPLWNTYSEQYIKWIFLAKFPKYPNVIGADINEPHGEAGWTGDER-DP 207

QY 179 RLAAERAGNAVLNVNPNLLIFVEGVQ-----SYNGDSYV--WGNLQAGQYPPVL 227  
Db 208 RLFAEKVGREILLKVAHLLIFVEGTQYTHVPNIDEIIEKKGWMTWFGENLGMVKDQPVRL 267

QY 228 NVPNRLVYSAHDYATSVGQTFWSDPTFPNNMNGIWNQWGLYFNQNIAPVWLGEFGTTL 287

Db 268 -PRGKVVYSPHVYGPSYVMDYFKSPDPFNNMPEIIEWHFGYLTDLNLYLV-IGEWGNY 325

QY 288 QSTTDQWLKTLVQYLRLPTAQYGADSFQWTFWNNPDSGDTGGLKDDWQTVTK 343  
Db 326 EG-LDKYWQDAFVKWLKIKKIY-----NFFYWCINPESGDTGGLFLDDWKTAVWEK 375

## RESULT 14

US-10-369-493-9473  
; Sequence 9473, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 9473  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Xylella fastidiosa  
US-10-369-493-9473

Query Match 33.2%; Score 660.5; DB 15; Length 369;  
Best Local Similarity 40.5%; Pred. No. 8.3e-56;  
Matches 143; Conservative 56; Mismatches 125; Indels 29; Gaps 12;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLNSRDYRMLDQIKSLGNTIRLP 64  
Db 26 YSISHGKVIDDKGN-QIQKGGSWFGFETTNVHVLWTRNWKFTIQSGMLNAVRLP 84

QY 65 YSDIILKPGTMPNSINFYQMNQDLOGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQS 124  
Db 85 FCPANLNSNTSPSID-YSRNPDQLQGLSLQILDKVVKLSDRRMVYLLDHRSDCSAIS 143

QY 125 ALWYTSVSEATWISDLQALARYKGNPTVGFDLNHEPHDPACWCGDPSIDWELAAER 184  
Db 144 ELWYDYSYSEKQWIDDLRFVAHRYANVPVIGLVDKNEPHGRATWGTGDKTDMNTAVEH 203

QY 185 AGNAVLNVNPNLLIFVEGVQSYNGDS-----YWGNGMLQAGQYPPVNLVNP-NRLVY 235  
Db 204 AAAAILEAAPKWLIGVEGI---GENPICSSTIGHFNGENLEPMDCTP--LKVPAHLLI 257

QY 236 SAHDYATSVGQTFWSDPTFPNNMNGIWNQWGLYFNQNIAPVWLGEFGTTL--QSTTDQ 293  
Db 258 MPHVGPDVYVQPYFNPSDPFPNMAAIDKHFGHFAKAGYA-MAIGFEGKYGEGDPRDI 316

QY 294 TLKTLVQYLRLPTAQYG-ADSFQWTFWNNPDSGDTGGLKDDWQTV--DTPK 343  
Db 317 AWQNAFVDYL---ISIGVTDAF---YWAANQNSGDTGGMVGNWDTTPRDDKVK 363

## RESULT 15

US-10-369-493-9262  
; Sequence 9262, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 9262  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Xylella fastidiosa  
US-10-369-493-9262

```
Db 167 YTEDFSEEDPINTWIEVAKRFKGMVNIAGDLKNEPHSVTSPAAAYTGTGTGATWGNPA 226
Qy 176 IDWLAERAGNAVLVSNPDLIFVGVQSYN-----GDSYWMGNLQAGQYVV 226
Db 227 TDWLAERIGKAILKVAPHWLIIFVGTQFTPKDSSYKGNVWGNLMAVXDYFV- 285
Qy 227 LNVP-NRLVSAHDYATSVGPOTWFSDDPT-FPNMFGIWNKNWGLFQNIAPVWLGERG 284
Db 286 -NLPRNKLVSPHYGPDVYNQYFPGPAKGFDPNLDIWHYHFGYVVKLELGYSVVIGFEG 344
Qy 285 TTL---QSTTDQTLKTLVQLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQTV 339
Db 345 GRYGHGDPDRVWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 10
US-10-369-493-1281
; Sequence 1281, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1281

Query Match 38.0%; Score 755; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.3e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDSYMLDQIKSLGYNTIRLPYS 67
Db 54 TSGBE-----TPHILFGVWFGFETPNHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFYQMNQDGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSQSALW 127
Db 108 ESKVPGTQFIGID-YXKNPDLRGLDSLOIMEKIKKAGDLGIFVLLDYHHRIGCTHIEPLW 166
Qy 128 YTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWGCSDPS 175
Db 167 YTEDFSEEDPINTWIEVAKRFKGMVNIAGDLKNEPHSVTSPAAAYTGTGTGATWGNPA 226
Qy 176 IDWLAERAGNAVLVSNPDLIFVGVQSYN-----GDSYWMGNLQAGQYVV 226
Db 227 TDWLAERIGKAILKVAPHWLIIFVGTQFTPKDSSYKGNVWGNLMAVXDYFV- 285
Qy 227 LNVP-NRLVSAHDYATSVGPOTWFSDDPT-FPNMFGIWNKNWGLFQNIAPVWLGERG 284
Db 286 -NLPRNKLVSPHYGPDVYNQYFPGPAKGFDPNLDIWHYHFGYVVKLELGYSVVIGFEG 344
Qy 285 TTL---QSTTDQTLKTLVQLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQTV 339
Db 345 GRYGHGDPDRVWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397

RESULT 11
US-10-369-493-20347
; Sequence 20347, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1281
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1281

Query Match 38.0%; Score 755; DB 15; Length 458;
Best Local Similarity 45.0%; Pred. No. 6.3e-65;
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

Qy 8 TSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDSYMLDQIKSLGYNTIRLPYS 67
Db 54 TSGBE-----TPHILFGVWFGFETPNHVHGLWKNWEDMLLQIKSLGFNAIRLPFCT 107
Qy 68 DILKPGTWPNSINFYQMNQDGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSQSALW 127
Db 108 ESKVPGTQFIGID-YXKNPDLRGLDSLOIMEKIKKAGDLGIFVLLDYHHRIGCTHIEPLW 166
Qy 128 YTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHD-----PACWGCSDPS 175
Db 167 YTEDFSEEDPINTWIEVAKRFKGMVNIAGDLKNEPHSVTSPAAAYTGTGTGATWGNPA 226
Qy 176 IDWLAERAGNAVLVSNPDLIFVGVQSYN-----GDSYWMGNLQAGQYVV 226
Db 227 TDWLAERIGKAILKVAPHWLIIFVGTQFTPKDSSYKGNVWGNLMAVXDYFV- 285
Qy 227 LNVP-NRLVSAHDYATSVGPOTWFSDDPT-FPNMFGIWNKNWGLFQNIAPVWLGERG 284
Db 286 -NLPRNKLVSPHYGPDVYNQYFPGPAKGFDPNLDIWHYHFGYVVKLELGYSVVIGFEG 344
Qy 285 TTL---QSTTDQTLKTLVQLRPTAQYGADSFQWTFWSWNPDSGDTGGILKDDQTV 339
Db 345 GRYGHGDPDRVWQNLVDWM--IENKFCDF--YWSWNPDSGDTGGILQDDWTTI 397
```

;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: US 60/279,493  
;; PRIOR FILING DATE: 2001-03-28  
;; NUMBER OF SEQ ID NOS: 69  
;; SEQ ID NO 50  
;; LENGTH: 535  
;; TYPE: PRT  
;; ORGANISM: Xanthomonas campestris  
US-09-927-827-50

Query Match 39.9%; Score 793.5; DB 10; Length 535;  
Best Local Similarity 46.9%; Pred. No. 1.3e-68;  
Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;  
  
QY 5 YWHTSGREILDANNVPRIAGINWFGFETCNVYVHGLMSRDYRSMDOIKSLGYNTIRLP 64  
DB 77 YSINNSRQIVDDSGKVQLKGVNVFETGNEVHGLMARNWKMIVQWQGLGFNAVRLP 136  
  
QY 65 YSDDIILKPTMPSINFYOMNODLOGLTSLOYMDKIVAYAGQIGRIILDRHRPDCSGS 124  
DB 137 FCPATLRSIMPASID-YSRNADLOGLTSLOLDKRVIAEFNARGMYVLDDHHTPCAGIS 195  
  
QY 125 ALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDMRLAAR 184  
DB 196 ELWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDMRKAAR 255  
  
QY 185 AGNAVLSVNPILLIFVEGVQ-----SYNGDSYVWGNLQAGQYPVNLNVP-NRLVYSAH 238  
DB 256 GSAAVLAVAPKWLIAVSGITDNPVCSNG-GIFWGNLQPLACTP--LNIPANRLLLAPH 312  
  
QY 239 DYATSVGPOTWFSDDPTFPNNMPCIMNKNWGLFNQNIAPVWLGEFGTTL--QSTTDQTL 296  
DB 313 VGPVDFVGSYFNDSFPNNMFAWERHFGQAGTH--ALLGEFDGKYGEGDARDKIQW 370  
  
QY 297 KTLVOYLRTAQYAGDSFQWTFWSNPDSDGTGGILKDDQTV 339  
DB 371 DALVKYLR---SKGIN--QGFYWSNPNDSGTGGILRDDWTSV 408

## RESULT 8

US-09-888-224-2  
;; Sequence 2, Application US/09888224  
;; Patent No. US20020120118A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Lam, D. et al.  
;; TITLE OF INVENTION: Endoglucanases  
;; NUMBER OF SEQUENCES: 6  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
;; ADDRESSEE: CECCHI, STEWART & OLSTEIN  
;; STREET: 6 BECKER FARM ROAD  
;; CITY: ROSELAND  
;; STATE: NEW JERSEY  
;; COUNTRY: USA  
;; ZIP: 07068  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 3.5 INCH DISKETTE  
;; COMPUTER: IBM PS/2  
;; OPERATING SYSTEM: MS-DOS  
;; SOFTWARE: WORD PERFECT 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/888,224  
;; FILING DATE:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/430,669  
;; FILING DATE: 28-Oct-1999  
;; APPLICATION NUMBER: US/08/651,572  
;; FILING DATE: 22-MAY-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Charles J. Herron  
;; REGISTRATION NUMBER: 28,019  
;; REFERENCE/DOCKET NUMBER: 331400-48  
;; TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 533 AMINO ACIDS  
;; TYPE: AMINO ACID  
;; STRANDEDNESS: Unknown  
;; TOPOLOGY: LINEAR  
;; MOLECULE TYPE: PROTEIN  
US-09-888-224-2

Query Match 38.3%; Score 762; DB 9; Length 553;  
Best Local Similarity 45.2%; Pred. No. 1.7e-65;  
Matches 160; Conservative 50; Mismatches 108; Indels 36; Gaps 11;  
  
QY 13 ILDANNVPRIAGINWFGFETCNVYVHGLMSRDYRSMDOIKSLGYNTIRLPYSDILKP 72  
DB 4 VATGEETPIILFGLWVWFGFETPNVYVHGLMSRWEDMLQIKSLGFNAIRLPCTQSVKP 63  
  
QY 73 GTMPSINFYOMNODLOGLTSLOYMDKIVAYAGQIGRIILDRHRPDCSQSALWTSSV 132  
DB 64 GTMPTAID-YAKNPDLOGLDSVQIMEXIIKKAGDLGIFVLDDYHRCNFIEPLWYDTSF 122  
  
QY 133 SEATWISDLOALAQRYKGNPTVVGFDLHNEPHDP-----ACWCGDPSIDWRL 180  
DB 123 SEQDIYINTWVEVAQRFGKYWNVIADLKNEPHSSPAPAAAYTDGSGATMCGNNAIDWNL 182  
  
QY 181 AAEAGNAVLSVNPILLIFVEGVQ-----SYN-GDSYVWGNLQAGQYPVNLNVP- 230  
DB 183 AAEISGAILLEVAQWVIFVEGTQFTTPEIDGRYKGNHNAWGNLNGVRYKPYV--NLPR 240  
  
QY 231 NRLVYSAHDYATSVGPOTWFSDDPTFPNNMPCIMNKNWGLFNQNIAPVWLGEFGTTL- 287  
DB 241 DKVYVSQVYGVSEYVDQYF-DPCEGFPDNLPELWVHFGYVVKLDLGYPVVIGFSGKYG 299  
  
QY 288 --QSTTDQTLKTLVOYLRTAQYAGDSFQWTFWSNPDSDGTGGILKDDQTV 339  
DB 300 HGGDPRDVTWQNKIIDWM--IQNKFCDF--YMSWNPNSGDTGGILKDDWTTI 348

## RESULT 9

US-10-293-344A-2  
;; Sequence 2, Application US/10293344A  
;; Publication No. US20030129723A1  
;; GENERAL INFORMATION:  
;; APPLICANT: RAKUTO KASEI INDUSTRIAL CO., LTD. and  
;; APPLICANT: National Institute of Advanced Industrial Science and Technology  
;; TITLE OF INVENTION: Thermophilic endoglucanase  
;; FILE REFERENCE: P017123A  
;; CURRENT APPLICATION NUMBER: US/10/293,344A  
;; CURRENT FILING DATE: 2003-03-17  
;; NUMBER OF SEQ ID NOS: 4  
;; SEQ ID NO 2  
;; LENGTH: 458  
;; TYPE: PRT  
;; ORGANISM: Pyrococcus horikoshii  
;; FEATURE:  
US-10-293-344A-2

Query Match 38.0%; Score 755; DB 14; Length 458;  
Best Local Similarity 45.0%; Pred. No. 6.3e-65;  
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;  
  
QY 8 TSGREILDANNVPRIAGINWFGFETCNVYVHGLMSRDYRSMDOIKSLGYNTIRLPYSD 67  
DB 54 TSCEE-----TPHLFGVNNWFGFETPNVYVHGLMKNWEDMLQIKSLGFNAIRLPFCT 107  
  
QY 68 DILKPGTWPNSINFYOMNODLOGLTSLOYMDKIVAYAGQIGRIILDRHRPDCSQSALW 127  
DB 108 ESVKFGTPIGID-YSKNPDLRGLDSLQIMEXIIKKAGDLGIFVLDDYHRCNFIEPLW 166  
  
QY 128 YTSSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHD-----PACWCGDPS 175

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRLP 64  
DB 26 YSINNSRQIVDDSGKVQLGKGVNVFGETGNHVMHGLWNRWMDMIYQVQGLGFNAVRLP 85  
QY 65 YSDDLKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGOS 124  
DB 86 FCPATLRSDTWPASID-YSRNADLQGLTSLQILDKVIAEFNARGMVLDDHHTPDCAGIS 144  
QY 125 ALWYTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184  
DB 145 ELWYTSYTEAQLADLRFAVNRKYNVPYVIGLDLKNEPHGAATWGTGNAATDWNKAAER 204  
QY 185 AGNVLISVNPULLIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVNLNVP-NRLVYSAH 238  
DB 205 GSAAVLAVAPKWLIAVEGIDTNPCVSTNG-GIFWGNLQPLACTP--LNIPANELLALPH 261  
QY 239 DYATSVGPQTWFSDDPTFPNNMFGIWNKNWGYLFNQNIAPIVWLGEFGTTL--QSTTDQTL 296  
DB 262 VYGPDVVQSYFNSDNPNNMFAIWERHFGQFAGTH--ALLLGEFDGKYGEGDARDKTWQ 319  
QY 297 KTLVQYLRPTAQYCADSFQWTFWSNPNDSGDTGGILKDDQTV 339  
DB 320 DALVKYLR---SKGIN--QGFYWSNPNNSGDTGGILRDDWTSV 357

## RESULT 5

US-10-369-493-15740  
; Sequence 15740, Application US/10369493  
; Publication No. US20030233675A1

## GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 15740  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
; NAME/KEY: unsure  
; LOCATION: (1)..(501)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-15740

Query Match 39.9%; Score 793.5; DB 15; Length 501;  
Best Local Similarity 46.9%; Pred. No. 1.2e-68;  
Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRLP 64  
DB 26 YSINNSRQIVDDSGKVQLGKGVNVFGETGNHVMHGLWNRWMDMIYQVQGLGFNAVRLP 85  
QY 65 YSDDLKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGOS 124  
DB 86 FCPATLRSDTWPASID-YSRNADLQGLTSLQILDKVIAEFNARGMVLDDHHTPDCAGIS 144  
QY 125 ALWYTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184  
DB 145 ELWYTSYTEAQLADLRFAVNRKYNVPYVIGLDLKNEPHGAATWGTGNAATDWNKAAER 204  
QY 185 AGNVLISVNPULLIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVNLNVP-NRLVYSAH 238  
DB 205 GSAAVLAVAPKWLIAVEGIDTNPCVSTNG-GIFWGNLQPLACTP--LNIPANELLALPH 261

QY 239 DYATSVGPQTWFSDDPTFPNNMFGIWNKNWGYLFNQNIAPIVWLGEFGTTL--QSTTDQTL 296  
DB 262 VYGPDVVQSYFNSDNPNNMFAIWERHFGQFAGTH--ALLLGEFDGKYGEGDARDKTWQ 319  
QY 297 KTLVQYLRPTAQYCADSFQWTFWSNPNDSGDTGGILKDDQTV 339  
DB 320 DALVKYLR---SKGIN--QGFYWSNPNNSGDTGGILRDDWTSV 357

## RESULT 6

US-10-369-493-15373  
; Sequence 15373, Application US/10369493  
; Publication No. US20030233675A1

## GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 15373  
; LENGTH: 518  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-15373

Query Match 39.9%; Score 793.5; DB 15; Length 518;  
Best Local Similarity 46.9%; Pred. No. 1.3e-68;  
Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;

QY 5 YWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNTIRLP 64  
DB 26 YSINNSRQIVDDSGKVQLGKGVNVFGETGNHVMHGLWNRWMDMIYQVQGLGFNAVRLP 85  
QY 65 YSDDLKPGTWPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGOS 124  
DB 86 FCPATLRSDTWPASID-YSRNADLQGLTSLQILDKVIAEFNARGMVLDDHHTPDCAGIS 144  
QY 125 ALWYTSVSSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAER 184  
DB 145 ELWYTSYTEAQLADLRFAVNRKYNVPYVIGLDLKNEPHGAATWGTGNAATDWNKAAER 204  
QY 185 AGNVLISVNPULLIFVEGVQ-----SYNGDSYWMGNLQAGQYPPVNLNVP-NRLVYSAH 238  
DB 205 GSAAVLAVAPKWLIAVEGIDTNPCVSTNG-GIFWGNLQPLACTP--LNIPANELLALPH 261  
QY 239 DYATSVGPQTWFSDDPTFPNNMFGIWNKNWGYLFNQNIAPIVWLGEFGTTL--QSTTDQTL 296  
DB 262 VYGPDVVQSYFNSDNPNNMFAIWERHFGQFAGTH--ALLLGEFDGKYGEGDARDKTWQ 319  
QY 297 KTLVQYLRPTAQYCADSFQWTFWSNPNDSGDTGGILKDDQTV 339  
DB 320 DALVKYLR---SKGIN--QGFYWSNPNNSGDTGGILRDDWTSV 357

## RESULT 7

US-09-927-827-50  
; Sequence 50, Application US/09927827  
; Publication No. US20030036176A1

## GENERAL INFORMATION:

; APPLICANT: Bower, Stanley G.  
; APPLICANT: Ramseier, Thomas M.  
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris  
; FILE REFERENCE: 38-10(15824)B  
; CURRENT APPLICATION NUMBER: US/09/927,827

QY 181 AAFRAGNAVLSVNPMLLIFVEGVQSYNGDSYWGNGNLOGAGQYPVLLNPNRLVYSAHDY 240  
Db 222 AAFRAGNAVLSVNPMLLIFVEGVQSYNGDSYWGNGNLOGAGQYPVLLNPNRLVYSAHDY 281  
QY 241 ATSPGQTFWSDPTFPNNMPPGIMWKNWGYLFNQNIAPVWLGEFGTTLOSTTDQWLKTLV 300  
Db 282 ATSPGQTFWSDPTFPNNMPPGIMWKNWGYLFNQNIAPVWLGEFGTTLOSTTDQWLKTLV 341  
QY 301 QVLRPTAQYAGDSFQWTFWSDPTFPNNMPPGIMWKNWGYLFNQNIAPVWLGEFGTTLOSTTDQWLKTLV 358  
Db 342 QVLRPTAQYAGDSFQWTFWSDPTFPNNMPPGIMWKNWGYLFNQNIAPVWLGEFGTTLOSTTDQWLKTLV 399

## RESULT 2

US-10-360-101-212  
; Sequence 212, Application US/10360101  
; Publication No. US20040009350A1  
; GENERAL INFORMATION:  
; APPLICANT: Moll, N.  
; APPLICANT: Leenhouts, Cornelis J.  
; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way  
; FILE REFERENCE: 2183-5673  
; CURRENT APPLICATION NUMBER: US/10/360,101  
; CURRENT FILING DATE: 2003-02-07  
; PRIOR APPLICATION NUMBER: EP 02077060.8  
; PRIOR FILING DATE: 2002-05-24  
; NUMBER OF SEQ ID NOS: 309  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 212  
; LENGTH: 397  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: sequence of cellulase  
US-10-360-101-212

Query Match 52.9%; Score 1052; DB 15; Length 397;  
Best Local Similarity 54.8%; Pred. No. 4.4e-94;  
Matches 194; Conservative 43; Mismatches 107; Indels 12; Gaps 4;  
QY 4 GYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNTRL 63  
Db 37 GYHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNTRL 96  
QY 64 PYSDDLKPGTWPNSINFPYQNNQDLQGLTSLOVMDKIVAYAGQIGLRIILDRHRPDCSGQ 123  
Db 97 PYSDDLKPGTWPNSINFPYQNNQDLQGLTSLOVMDKIVAYAGQIGLRIILDRHRPDCSGQ 155  
QY 124 SALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 183  
Db 156 SELWYTSQYPSERWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 215  
QY 184 RAGNAVLSVNPMLLIFVEGVQ-----SYNGDSYWGNGNLOGAGQYPVLLNPNRLVYSAHDY 240  
Db 216 RAGNAVLSVNPMLLIFVEGVQ-----SYNGDSYWGNGNLOGAGQYPVLLNPNRLVYSAHDY 275  
QY 241 ATSPGQTFWSDPTFPNNMPPGIMWKNWGYLFNQNIAPVWLGEFG--TTLOSTTDQWLKTL 298  
Db 276 GPGVSSQPFNDPAPNSLPALWDQWGLYLSKQNIAPVWLGEFGGRNVDLSCPEGKQWA 335  
QY 299 LVQYLRPTAQYAGDSFQWTFWSDPTFPNNMPPGIMWKNWGYLFNQNIAPVWLGEFGTTLOSTTDQWLKTLV 354  
Db 336 LVHYI-----GANNLYFTYWSLNPNNSGDTGGILKDDQVTVDKGYLAPIKSSIFDPV 385

## RESULT 3

US-10-369-493-16379  
; Sequence 16379, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B/US/10/369,493  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 16379  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-16379

Query Match 40.3%; Score 800.5; DB 15; Length 493;  
Best Local Similarity 47.2%; Pred. No. 2.5e-69;  
Matches 162; Conservative 58; Mismatches 104; Indels 19; Gaps 9;  
QY 5 YHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNTRL 64  
Db 26 YHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYRMLDQIKSLGYNTRL 85  
QY 65 YSDDLKPGTWPNSINFPYQNNQDLQGLTSLOVMDKIVAYAGQIGLRIILDRHRPDCSGQ 124  
Db 86 YSDDLKPGTWPNSINFPYQNNQDLQGLTSLOVMDKIVAYAGQIGLRIILDRHRPDCSGQ 144  
QY 125 ALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 184  
Db 145 ELWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRLAAE 204  
QY 185 AGNAVLSVNPMLLIFVEGVQ-----SYNGDSYWGNGNLOGAGQYPVLLNPNRLVYSAH 238  
Db 205 GSAAVLAVAPKWLIAVEGTTDNPVCSTNG-GIFWGNLQPLACTP--LNIPIALLLAPH 261  
QY 239 DYATSVGPQTFWSDPTFPNNMPPGIMWKNWGYLFNQNIAPVWLGEFGTTL--QSTTDQWL 296  
Db 262 VYGPVDFVQSYFNDSPNNPNNPALWERHFGQFAGTH--ALLGFEFGKYGEGDARDKTWQ 319  
QY 297 KTLVQYLRPTAQYAGDSFQWTFWSDPTFPNNMPPGIMWKNWGYLFNQNIAPVWLGEFGTTL--QSTTDQWL 339  
Db 320 DALVYLR-----SKGIN--QGFYWSNPNNSGDTGGILKDDQVTV 357

## RESULT 4

US-10-369-493-16123  
; Sequence 16123, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 16123  
; LENGTH: 483  
; TYPE: PRT  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-16123

Query Match 39.9%; Score 793.5; DB 15; Length 483;  
Best Local Similarity 46.9%; Pred. No. 1.2e-68;  
Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: August 2, 2004, 16:34:32 ; Search time 41.3333 Seconds  
(without alignments)  
2716.899 Million cell updates/sec

Title: US-09-997-504A-10  
Perfect score: 1987  
Sequence: 1 AGGYWHTSGREILDANNVP.....VDTKDYGIAPKSSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1291235 segs, 313682936 residues  
Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09D\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US10F\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1978	99.5	562	9	US-09-981-900B-5
2	1052	52.9	397	15	US-10-360-101-212
3	800.5	40.3	493	15	US-10-369-493-16379
4	793.5	39.9	483	15	US-10-369-493-16123
5	793.5	39.9	501	15	US-10-369-493-15740
6	793.5	39.9	518	15	US-10-369-493-15373
7	793.5	39.9	535	15	US-09-927-827-50
8	762	38.3	553	9	US-09-988-224-2
9	755	38.0	458	14	US-10-293-344A-2
10	755	38.0	458	15	US-10-369-493-1281
11	755	38.0	458	15	US-10-369-493-20347
12	687.5	34.6	592	15	US-10-369-493-17534
13	682.5	34.3	514	15	US-10-369-493-21618
14	660.5	33.2	369	15	US-10-369-493-9473
15	655.5	33.0	421	15	US-10-369-493-9262

16	582	29.3	472	15	US-10-369-493-8139	Sequence 8139, Ap
17	453.5	22.8	362	15	US-10-369-493-3080	Sequence 3080, Ap
18	276.5	13.9	353	15	US-10-369-493-15733	Sequence 15733, A
19	276.5	13.9	353	15	US-10-369-493-16119	Sequence 16119, A
20	276.5	13.9	582	10	US-09-927-827-48	Sequence 48, Appl
21	257.5	13.0	533	12	US-10-424-599-176232	Sequence 176232, A
22	244	12.3	375	15	US-10-369-493-12597	Sequence 12597, A
23	234	11.8	397	15	US-10-369-493-3152	Sequence 3152, Ap
24	191.5	9.6	438	12	US-10-425-114-66753	Sequence 66753, A
25	190.5	9.6	314	15	US-10-369-493-15726	Sequence 15726, A
26	190.5	9.6	314	15	US-10-369-493-16112	Sequence 16112, A
27	190.5	9.6	337	15	US-10-369-493-15359	Sequence 15359, A
28	189.5	9.5	917	16	US-10-437-963-181484	Sequence 181484, A
29	188	9.5	644	16	US-10-437-963-139883	Sequence 139883, A
30	185	9.3	517	12	US-10-107-431-259	Sequence 259, App
31	173.5	8.7	375	10	US-09-917-378-3	Sequence 3, Appli
32	173.5	8.7	375	10	US-09-917-378-6	Sequence 6, Appli
33	173.5	8.7	762	10	US-09-917-378-1	Sequence 1, Appli
34	168.5	8.5	346	10	US-09-917-378-8	Sequence 8, Appli
35	168	8.5	356	10	US-09-917-378-7	Sequence 7, Appli
36	166.5	8.4	558	16	US-10-437-963-185049	Sequence 185049, A
37	160.5	8.1	722	12	US-10-282-122A-51993	Sequence 51993, A
38	156.5	7.9	518	10	US-09-769-734-56	Sequence 56, Appl
39	155.5	7.8	356	15	US-10-369-493-17756	Sequence 17756, A
40	153.5	7.7	1449	15	US-10-437-963-176100	Sequence 176100, A
41	152.5	7.7	533	12	US-10-424-599-150988	Sequence 150988, A
42	150.5	7.6	268	12	US-10-424-599-284543	Sequence 284543, A
43	150.5	7.6	315	15	US-10-369-493-9177	Sequence 9177, Ap
44	147	7.4	217	16	US-10-437-963-112718	Sequence 112718, A
45	146	7.3	313	15	US-10-369-493-15674	Sequence 15674, A

ALIGNMENTS

RESULT 1  
US-09-981-900B-5  
; Sequence 5, Application US/09981900B  
; Patent No. US20020138878A1  
; GENERAL INFORMATION:  
; APPLICANT: Sticklen, Masomeh B  
; APPLICANT: Magbool, Shahina B  
; APPLICANT: Dale, Bruce E  
; TITLE OF INVENTION: TRANSGENIC PLANTS CONTAINING LIGNINASE AND CELLULOSE TO FERMENTABLE SUGARS  
; TITLE OF INVENTION: AND CELLULOSE TO FERMENTABLE SUGARS  
; FILE REFERENCE: MSU 4.1-539  
; CURRENT APPLICATION NUMBER: US/05/981,900B  
; CURRENT FILING DATE: 2002-03-18  
; PRIOR FILING DATE: 2002-03-18  
; PRIOR FILING DATE: 2000-10-20  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 562  
; TYPE: PRT  
; ORGANISM: Acidothermus cellulolyticus  
US-09-981-900B-5

Query Match	99.5%	Score	1978	DB	9	Length	562
Best Local Similarity	99.7%	Pred	No. 1,6e-184				
Matches	557	Conservative	0	Mismatches	1	Indels	0
Gaps	0						
QY	1	AGGYWHTSGREILDANNVPVRIAGINWFGPFCNYYVHGLWSRDYRSMDOIKSLGYNT	60				
DB	42	AGGYWHTSGREILDANNVPVRIAGINWFGPFCNYYVHGLWSRDYRSMDOIKSLGYNT	101				
QY	61	IRLPYSDDLKPGTNPNSINFYQNDLQGLTSQWMDKIYVAGQIGRLIRLDRHPDC	120				
DB	102	IRLPYSDDLKPGTNPNSINFYQNDLQGLTSQWMDKIYVAGQIGRLIRLDRHPDC	161				
QY	121	SGQSALWYTSVSEATWISDIQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL	180				
DB	162	SGQSALWYTSVSEATWISDIQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL	221				





US-08-525-697-2  
 ; Sequence 2, Application US/08525697  
 ; Patent No. 5795764  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Christgau, Stephan  
 ; APPLICANT: Andersen, Lene N  
 ; APPLICANT: Kauppinen, Sakari  
 ; APPLICANT: Helst-Hansen, Hans P  
 ; APPLICANT: Dalboege, Henrik  
 ; TITLE OF INVENTION: AN ENZYME EXHIBITING MANNANASE ACTIVITY  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; STREET: No. 57957640 No. 5795764disk of No. 5795764th America, Inc.  
 ; ADDRESS: 405 Lexington Avenue, 64th Floor  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: United States of America  
 ; ZIP: 10174-6401  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: 21-SEP-1995  
 ; APPLICATION NUMBER: US/08/525,697  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Harrington, James J.  
 ; REGISTRATION NUMBER: 38,711  
 ; REFERENCE/DOCKET NUMBER: 4004.204-US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 212-867-0123  
 ; TELEFAX: 212-878-9655  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: DK 0486/93  
 ; FILING DATE: 30-APR-1993  
 ; CLASSIFICATION: 435  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 377 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; US-08-525-697-2

Query Match 8.5%; Score 168.5; DB 1; Length 377;  
 Best Local Similarity 23.1%; Pred. No. 5.4e-08;  
 Matches 82; Conservative 56; Mismatches 116; Indels 101; Gaps 21;

Qy 24 AGIN--WFGETCTVYVHGLWSRDYRSMQLDKSLGYNTIRLPSYSDILKPGTTPNSINF 81  
 Db 50 AGTNSYWGIFLTNNDDVDLWNSQLAASDLKILRVWGFNDVNTKPTD-----W 99

Qy 82 YQMCD-----LQGLTSIQWMDKIVAYAGQIGLRILD--RHRPDCSGQSAL----- 126  
 Db 100 YQLHANGTSINTGADGLQRLDYVTSAEKYGVKLIINFVNEWTYDGGMOAYVYAYGAA 159

Qy 127 ---WYTSVSEATWISDLQALQRYKGNPTVGFDLNHEPHDPACGCGDPSIDWLAEE 183  
 Db 160 QTDFTYNTAQAYKNYIKAVWSYSSAAIFAWELANEPR---COGC-DTSVLYNWSID 215

Qy 184 RAGNAVLNVNPLLIIFVEGVQSYNGDSYWGNGNLQAGQYPVV---LNVNPLRVYSADH 239  
 Db 216 -TSKYIKLSLKHVLT-----GDE-GFGLVDSDGSYPYTYGSLNFTKLGISTID 266

Qy 240 YAT-SVGPQWFSFPTFPNMPGHWKNGYLFNONTA-----PWLGEFGTT----- 286  
 Db 267 FGLTLHPDSNGTSYD-----WNGW---ITAHAAACAVGKPCLLSEYGVTSNHCA 315

Qy 287 LQSTTDOT-----WLKTLVOYLRPTAQYGADSFQWTF-KSNWNPDSGDT 328  
 Db 316 VESPWQQTAGNATGISGLYH-----QYGT-----TFSKQSPNDGNT 353

RESULT 13  
 US-09-797-464A-4  
 ; Sequence 4, Application US/09797464A  
 ; Patent No. 6630340  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wiltting, Reinhard  
 ; APPLICANT: Bjornvad, Mads Eskelund  
 ; APPLICANT: Kauppinen, Markus Sakari  
 ; APPLICANT: Schulein, Martin  
 ; TITLE OF INVENTION: Family 5 Xyloglucanases  
 ; FILE REFERENCE: 6073.200-US  
 ; CURRENT APPLICATION NUMBER: US/09/797,464A  
 ; CURRENT FILING DATE: 2002-02-19  
 ; NUMBER OF SEQ ID NOS: 25  
 ; SOFTWARE: Patentin version 3.1  
 ; SEQ ID NO 4  
 ; LENGTH: 400  
 ; TYPE: PRT  
 ; ORGANISM: Paenibacillus sp.  
 ; US-09-797-464A-4

Query Match 7.2%; Score 143; DB 4; Length 400;  
 Best Local Similarity 22.5%; Pred. No. 1.8e-05;  
 Matches 86; Conservative 53; Mismatches 131; Indels 112; Gaps 23;

Qy 3 GGYWHTSGREILDANNVFPRIAGINWFGFETCTNYVHGLWSRDYRSMQLDKSLGYNTIR 62  
 Db 48 GAGWNLGNQLEATYNGVPSETA---W-----GNPVV-----TPELIKVKVKAAGFKTIR 92

Qy 63 LP--YSDDLKPGTNP--SINFQYQNNQDLQGLTSIQWMDKIVAYAGQIGLRILDHRPD 119  
 Db 93 IPVSYLNIH---GSAPNTYINAALNR-----VQTVVDYAYNEGLVYVINIRGDD 139

Qy 120 CSGQSALWYTSVSEATWISD---LQALAORYKGNPTVGVFDLHNEPHDPACWCGDGP 174  
 Db 140 YNSIFGGWLLVNGSNQAAIKEKYQKVQOIATKSNYNERLIFESMNEVFD---GNVGNP 196

Qy 175 SIDWELAAERAGNAVLNVNPL--IFVEGVQSY--NODSYW-----WGNLQ-GAQYYP 224  
 Db 197 -----NAAYYANLNAVYQIFVDYVTRQGGNNARWLLIPGWNTNIDYTVGNYG 244

Qy 225 VVL-----NVEN---RLVVSADY-----ATSVGP---QTVFSDP 253  
 Db 245 FALPTDHFSSAIPSSQKRIWISAHYSPWDFAGEENGNIQWGAATNPSSKSTWQGED 304

Qy 254 TFPNNMPGIWKK--NWGYLFNQNIAPVWLGEFGT---TLQSTTDQTLKLVQYLRPTA- 307  
 Db 305 YLNAQFKSMYDKFVTQGY-----PVVIGBFGSIDKTAIDSTNNVYRQAYAKAVTATAK 357

Qy 308 QYGADSFQWTFWSNPNDSGDTG 329  
 Db 358 KYGAVPVTVW-----DNGHNG 372

RESULT 14  
 US-09-797-464A-7  
 ; Sequence 7, Application US/09797464A  
 ; Patent No. 6630340  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wiltting, Reinhard  
 ; APPLICANT: Bjornvad, Mads Eskelund  
 ; APPLICANT: Kauppinen, Markus Sakari  
 ; APPLICANT: Schulein, Martin  
 ; TITLE OF INVENTION: Family 5 Xyloglucanases  
 ; FILE REFERENCE: 6073.200-US  
 ; CURRENT APPLICATION NUMBER: US/09/797,464A  
 ; CURRENT FILING DATE: 2002-02-19  
 ; NUMBER OF SEQ ID NOS: 25  
 ; SOFTWARE: Patentin version 3.1  
 ; SEQ ID NO 7  
 ; LENGTH: 348  
 ; TYPE: PRT

QY 338 TVDTVKDGYLAP 349  
Db 562 TWDEQKYNFLKP 573

## RESULT 10

US-09-136-574A-43  
; Sequence 43, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.  
; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.

TITLE OF INVENTION: Compositions and Methods for  
Treating Cellulose Containing Fabrics Using Truncated  
Cellulase Enzyme Compositions

NUMBER OF SEQUENCES: 49

CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Center, P.O. Box 457  
CITY: Spring House  
STATE: PA  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/136,574A  
FILING DATE: 19-Aug-1998

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/932,571  
FILING DATE: September 19, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: 1997US001/CIP

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818

TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1426 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: No. 6294366e

SEQUENCE DESCRIPTION: SEQ ID NO: 43:

US-09-136-574A-43

Query Match 31.3%; Score 621; DB 3; Length 1426;  
Best Local Similarity 35.5%; Pred. No. 4.8e-51;  
Matches 132; Conservative 70; Mismatches 128; Indels 42; Gaps 11;

QY 10 GREILDANNVPRIAGINWFGFTCNVYVHGLSRDYSMLDQIKSLGYNTIRLPYSDDI 69  
Db 1022 GNKVIDKGFVMTLVGNVFGFTGTVFVGVNSCLKSLALIANRGNLLEVPISAEI 1081  
QY 70 L---KPTMPN--GINPYQMNDIQLGTSLOVMKIVAYAGQIGLRIILDRH--RPCSGQ 123  
Db 1082 ILNWSKGIYKPKNINY--VNPLEGLTSLEVFDFVVKTCVEGLKIMDIHSAKTAMGH 1140  
QY 124 -SALWYTSSEATWISDLQALQRYKGNVTYVGFDLHNEPHD-----ACWGCDFPS 175  
Db 1141 IYPVWYTDITPDDYKACEWITERYKNDDTIIVAFDLKNEPHGKPHQDVSFAKWDNSTDI 1200

QY 176 IDWELAERAGNAVLISUNENLLIFVSGVQSYNGD-----SYWVGGLQAGQ 222  
Db 1201 NNMYAAETCAKRIILAKNPNLVIIEIAYPKDDVTWTSKSSSDYYSIWVGGLRGVK 1260  
QY 223 YPVVL-NVPNRLVYSADYATSYGPQTW-----FSDPTFPNMPGINWKNWGYLFNQNIAP 277  
Db 1261 YPINLQYQNKVVYSPEHYGFLVYQQFVFPFGFTKDTLYND---CWRDNWYIMDNGIAP 1317  
QY 278 VWLGEFGTTILOSITDQWTKTLVQYLRPTAQYGADSFQWTFMSWNPDSGDTGGILKDDWQ 337  
Db 1318 LLIGEWGGYLDGGNEKWMYLRDYII-----ENHIHFTFCYNANSGDTGLVGYDFS 1371  
QY 338 TVDTVKDGYLAP 349  
Db 1372 TWDEQKYNFLKP 1383

## RESULT 11

US-08-276-213-2

; Sequence 2, Application US/08276213

; Patent No. 5536655

; GENERAL INFORMATION:

; APPLICANT: Thomas, Steven

; APPLICANT: Laymon, Robert

; APPLICANT: Himmel, Michael

; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: National Renewable Energy Laboratory

; STREET: 1617 Cole Boulevard

; CITY: Golden

; STATE: CO

; COUNTRY: USA

; ZIP: 80401-3393

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/276,213

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: O'Connor, Edna

; REGISTRATION NUMBER: 29,252

; REFERENCE/DOCKET NUMBER: NREL IR# 94-08

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303)231-1000

; TELEFAX: (303)231-1098

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 38 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: N-terminal

US-08-276-213-2

Query Match 10.3%; Score 205; DB 1; Length 38;  
Best Local Similarity 97.4%; Pred. No. 4.9e-13;  
Matches 37; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGQWHTSGREILDANNVPRIAGINWFGFTCNVYV 38  
Db 1 AGGQWHTSGREILDANNVPRIAGINWFGFTCNVYV 38

## RESULT 12

Db 199 TDMNLAERIGKALIKVAPHLIFVEGTQFTNPKTSSYKMGYNAMWGNLMAVKDYPV- 257  
Qy 227 LNVP-NRLVYSAHDYATSVGPQWFSDEPT-FPNNMFGIMNKNMGYLFNQNIAPVWLGEGF 284  
Db 258 -NLPNKLIVSPHYGPDVYNQPYFGPAKGFDPNLPDIWYHFGYKLELGYSVWIGEGF 316  
Qy 285 TTL---QSTTDQWKLTVQVLRPTAQYAGDSQWTFWNSNPDGDTGGILKDDQWTV 339  
Db 317 GYGHGGPRDVIWQNLVDWM--IENKFCDF--YWSNPNPDGDTGGILQDDWTTI 369

## RESULT 8

US-09-869-197-5  
; Sequence 5, Application US/09869197  
; Patent No. 6566113  
; GENERAL INFORMATION:  
; APPLICANT: TAKAYAMA, Masanori  
; APPLICANT: UMEDA, Kahoko  
; APPLICANT: KAYAMA, No. 6566113uto  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikuroshin  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: TAKAYAMA-6  
; CURRENT APPLICATION NUMBER: US/09/869,197  
; CURRENT FILING DATE: 2001-06-25  
; PRIOR APPLICATION NUMBER: PCT/JP99/07009  
; PRIOR FILING DATE: 1999-12-14  
; PRIOR APPLICATION NUMBER: JP 366237/1998  
; PRIOR FILING DATE: 1998-12-24  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 458  
; TYPE: PRT  
; ORGANISM: Pyrococcus horikoshii OT3  
US-09-869-197-5

Query Match 38.0%; Score 755; DB 4; Length 458;  
Best Local Similarity 45.0%; Pred. No. 8.5e-65;  
Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;  
Qy 8 TSGREILDANNVPRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNTIRLPYSD 67  
Db 54 TSGEE-----TPIHLFGVNWFGFETENYVHGLWKENWEDMLLIQIKSLGFNAIRLPFCT 107  
Qy 68 DILKPGTNPNSINFYQNMQLGLTSLQVMDKIVAYAGQIGLRIILDRHRPDCSGQSALW 127  
Db 108 ESKVPGTQPIGID-YSKNPDURGDSLQIKWELIKKAGDGLIFVLDDYHRIKGTHTIEPLW 166  
Qy 128 YTSSVSSEATWISDLQALAQRYKGNPTVYVGFDLHNEPHD-----PACWCGCDPS 175  
Db 167 YTEDFSEEDFINTWIEVAKRFGKWNVIGADLKNPHSVTSPPAAVTDGTGATWGNPNA 226  
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----GDSYWGNGNLQAGQYPPV 226  
Db 227 TDMNLAERIGKALIKVAPHLIFVEGTQFTNPKTSSYKMGYNAMWGNLMAVKDYPV- 285  
Qy 227 LNVP-NRLVYSAHDYATSVGPQWFSDEPT-FPNNMFGIMNKNMGYLFNQNIAPVWLGEGF 284  
Db 286 -NLPNKLIVSPHYGPDVYNQPYFGPAKGFDPNLPDIWYHFGYKLELGYSVWIGEGF 344  
Qy 285 TTL---QSTTDQWKLTVQVLRPTAQYAGDSQWTFWNSNPDGDTGGILKDDQWTV 339  
Db 345 GYGHGGPRDVIWQNLVDWM--IENKFCDF--YWSNPNPDGDTGGILQDDWTTI 397

## RESULT 9

US-09-136-574A-47  
; Sequence 47, Application US/09136574A  
; Patent No. 6294366  
; GENERAL INFORMATION:  
; APPLICANT: Farrington, Graham K.

; Anderson, Paige  
; Gibbs, Moreland  
; Bergquist, Peter  
; Daniels, Roy  
; Morgan, Hugh W.  
; Williams, Diane P.  
; TITLE OF INVENTION: Compositions and Methods for  
; Treating Cellulose Containing Fabrics Using Truncated  
; Cellulase Enzyme Compositions  
; NUMBER OF SEQUENCES: 49  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Center, P.O. Box 457  
; CITY: Spring House  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/136,574A  
; FILING DATE: 19-Aug-1998  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/932,571  
; FILING DATE: September 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: 1997US001/CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 47:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 616 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 47:  
US-09-136-574A-47

Query Match 31.3%; Score 621; DB 3; Length 616;  
Best Local Similarity 35.5%; Pred. No. 1.3e-51;  
Matches 132; Conservative 70; Mismatches 129; Indels 42; Gaps 11;  
Qy 10 GREILDANNVPRIAGINWFGFETCNVYVHGLWSDYRSMLDQIKSLGYNTIRLPYSDDI 69  
Db 212 GNKIVDKGKPVWLTGVNFGFNTGVNFDGVWSCNLKALAEIANRGNLRLVPISAEI 271  
Qy 70 L---KPGTMEN-SINFYQNMQLGLTSLQVMDKIVAYAGQIGLRIILDRH--RPDCSGQ 123  
Db 272 ILNWSKGIYKPNINYY-VNPELGITSLVDFVVKTKVGLKIMLDIHSAXTDAMGH 330  
Qy 124 -SALWYTSVSEATWISDLQALAQRYKGNPTVYVGFDLHNEPHD-----ACWCGCDPS 175  
Db 331 IYVWYTTITPEDYKACEMITERYKNDITIVAFDLKNEPHGKFPWQDSVFAKWDNSTDI 390  
Qy 176 IDWELAAERAGNAVLNPNLLIFVEGVQSYN-----SYWNGNQLQAGQ 222  
Db 391 NNWKAATCAKRIILAKNPNMLIVIEGIEAPKDDVTWTSKSSDYSTWGGNLRGVK 450  
Qy 223 YPVVL-NVFNRLVYSAHDYATSVGPQW---FSDPTFPNNMFGIMNKNMGYLFNQNIAP 277  
Db 451 YPINLGOYQNVVSPHYGPDVYQPYFGTFTKDTLYND---CWRDNWYIINDNGIAP 507  
Qy 278 VWLGEFTITLQSTTDQWKLTVQVLRPTAQYAGDSQWTFWNSNPDGDTGGILKDDQW 337  
Db 508 LLIGEWGGLDGGDNKWKWYLRDYII-----ENHIIHHTFWCYNANSNGDTGGILVGYDFS 561





Query Match 99.5%; Score 1978; DB 1; Length 358;  
Best Local Similarity 99.7%; Pred. No. 1.1e-183;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60

QY 61 IRLPYSDIILKPGTMPNSINFYQMNODLQGLTSLQVMDKIVAYAGIGRIILDRHRPDC 120  
DB 61 IRLPYSDIILKPGTMPNSINFYQMNODLQGLTSLQVMDKIVAYAGIGRIILDRHRPDC 120

QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPISIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPISIDWRL 180

QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOGAGQYVVLNPNRLVYSAHDY 240  
DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOGAGQYVVLNPNRLVYSAHDY 240

QY 241 ATSVGPQTWFSDDPTFPNNMFGIWNKNWGYLFNONTAPWLGEFGTTLOSTTDDOTWTKTLV 300  
DB 241 ATSVGPQTWFSDDPTFPNNMFGIWNKNWGYLFNONTAPWLGEFGTTLOSTTDDOTWTKTLV 300

QY 301 QYLRPTAQYGADSFQMTFWSWNPDSGDTGGILKDDMQTVDTVKDGYLAPIKSSIFDPV 358  
DB 301 QYLRPTAQYGADSFQMTFWSWNPDSGDTGGILKDDMQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 2  
US-08-276-213-3  
; Sequence 3, Application US/08276213  
; Patent No. 5536655  
; GENERAL INFORMATION:  
; APPLICANT: Thomas, Steven  
; APPLICANT: Laymon, Robert  
; APPLICANT: Himmel, Michael  
; TITLE OF INVENTION: GENE ENCODING FOR THE E1 ENDOGLUCANASE  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: National Renewable Energy Laboratory  
; STREET: 1617 Cole Boulevard  
; CITY: Golden  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80401-3393  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: O'Connor, Edna  
; REGISTRATION NUMBER: 29,252  
; REFERENCE/DOCKET NUMBER: NREL IR# 94-08  
; TELEPHONE: (303)231-1000  
; TELEFAX: (303)231-1098  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 521 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; US-08-276-213-3

Query Match 99.5%; Score 1978; DB 1; Length 521;  
Best Local Similarity 99.7%; Pred. No. 2e-183;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMLDQIKSLGYNT 60

QY 61 IRLPYSDIILKPGTMPNSINFYQMNODLQGLTSLQVMDKIVAYAGIGRIILDRHRPDC 120  
DB 61 IRLPYSDIILKPGTMPNSINFYQMNODLQGLTSLQVMDKIVAYAGIGRIILDRHRPDC 120

QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPISIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGGDPISIDWRL 180

QY 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOGAGQYVVLNPNRLVYSAHDY 240  
DB 181 AERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLOGAGQYVVLNPNRLVYSAHDY 240

QY 241 ATSVGPQTWFSDDPTFPNNMFGIWNKNWGYLFNONTAPWLGEFGTTLOSTTDDOTWTKTLV 300  
DB 241 ATSVGPQTWFSDDPTFPNNMFGIWNKNWGYLFNONTAPWLGEFGTTLOSTTDDOTWTKTLV 300

QY 301 QYLRPTAQYGADSFQMTFWSWNPDSGDTGGILKDDMQTVDTVKDGYLAPIKSSIFDPV 358  
DB 301 QYLRPTAQYGADSFQMTFWSWNPDSGDTGGILKDDMQTVDTVKDGYLAPIKSSIFDPV 358

RESULT 3  
US-08-651-572-2  
; Sequence 2, Application US/08651572  
; Patent No. 5789228  
; GENERAL INFORMATION:  
; APPLICANT: Lam, D. et al.  
; TITLE OF INVENTION: Endoglucanases  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
; STREET: 6 BECKER FARM ROAD  
; CITY: ROSELAND  
; STATE: NEW JERSEY  
; COUNTRY: USA  
; ZIP: 07068  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 INCH DISKETTE  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: MS-DOS  
; SOFTWARE: WORD PERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/651,572  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Charles J. Heiron  
; REGISTRATION NUMBER: 28,019  
; REFERENCE/DOCKET NUMBER: 331400-48  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201-994-1700  
; TELEFAX: 201-994-1744  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 AMINO ACIDS  
; TYPE: AMINO ACID  
; STRANDEDNESS:  
; TOPOLOGY: LINEAR  
; MOLECULE TYPE: PROTEIN  
; US-08-651-572-2

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:31:56 ; Search time 14.6667 Seconds  
(without alignments)  
1260.143 Million cell updates/sec

Title: US-09-997-504A-10

Perfect score: 1987  
Sequence: 1 AGGYWHTSGREILDANNVP.....VDTVKGGLYAPIKSIFDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1978	99.5	358	1	US-08-604-913B-11
2	1978	99.5	521	1	US-08-276-213-3
3	762	38.3	553	1	US-08-651-572-2
4	762	38.3	553	3	US-09-066-544-2
5	762	38.3	553	3	US-08-951-086-2
6	762	38.3	553	4	US-09-430-669-2
7	755	38.0	430	4	US-08-869-197-1
8	755	38.0	458	4	US-09-889-197-5
9	621	31.3	616	3	US-09-136-574A-47
10	621	31.3	1426	3	US-08-136-574A-43
11	205	10.3	38	1	US-08-276-213-2
12	168.5	8.5	377	1	US-08-525-697-2
13	143	7.2	400	4	US-09-797-464A-4
14	139	7.0	348	4	US-09-797-464A-7
15	139	7.0	476	4	US-09-339-159B-4
16	139	7.0	490	4	US-09-339-159B-2
17	138	6.9	395	4	US-09-797-464A-2
18	136	6.8	363	4	US-09-797-464A-11
19	136	6.8	468	4	US-09-485-648-4
20	136	6.8	468	4	US-09-503-565-4
21	136	6.8	468	4	US-09-485-649-4
22	136	6.8	468	4	US-09-339-159B-8
23	136	6.8	493	4	US-09-485-648-2
24	136	6.8	493	4	US-09-503-565-2
25	136	6.8	493	4	US-09-485-649-2
26	136	6.8	493	4	US-09-339-159B-6
27	131	6.6	24	1	US-08-276-213-1

28	131	6.6	331	4	US-09-339-159B-12
29	127	6.4	331	3	US-08-849-751-4
30	127	6.4	331	4	US-08-478-816-4
31	122	6.1	320	4	US-09-339-159B-22
32	121.5	6.1	663	4	US-09-134-078-61
33	121.5	6.1	680	4	US-08-134-078-25
34	116.5	5.9	429	1	US-08-745-977-4
35	116.5	5.9	429	3	US-09-040-699A-4
36	114.5	5.8	360	4	US-09-134-078-27
37	113	5.7	317	2	US-08-066-075-2
38	113	5.7	317	3	US-08-518-615A-2
39	113	5.7	317	3	US-08-951-889-2
40	113	5.7	317	3	US-08-472-857-2
41	112.5	5.7	461	1	US-08-672-571A-3
42	112.5	5.7	490	1	US-08-672-571A-1
43	112	5.6	327	2	US-08-169-948B-16
44	112	5.6	327	2	US-08-448-873-16
45	112	5.6	327	3	US-08-382-452D-16

## ALIGNMENTS

RESULT 1  
US-08-604-913B-11  
; Sequence 11, Application US/08604913B  
; Patent No. 5712142  
; GENERAL INFORMATION:  
; APPLICANT: Adney, William S.  
; APPLICANT: Thomas, Steven R.  
; APPLICANT: Himmel, Michael E.  
; APPLICANT: Baker, John O.  
; APPLICANT: Chou, Yat-Chen  
; TITLE OF INVENTION: METHOD FOR INCREASING  
; THERMOSTABILITY IN CELLULOSE ENZYMES  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: National Renewable Energy Laboratory  
; STREET: 1617 Cole Boulevard  
; CITY: Golden  
; STATE: CO  
; COUNTRY: U.S.A.  
; ZIP: 80401-3393  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASC II (DOS) text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/604,913B  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/276,213  
; FILING DATE: 15-070-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Edna M. O'Connor  
; REGISTRATION NUMBER: 29,252  
; REFERENCE/DOCKET NUMBER: 95-56  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 303/384-7573  
; TELEFAX: 303/384-7499  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 358 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; FEATURE:  
; NAME/KEY: EI-CAT  
US-08-604-913B-11

Sequence 12, Appl  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 22, Appl  
Sequence 61, Appl  
Sequence 25, Appl  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 27, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 1, Appli  
Sequence 16, Appl  
Sequence 16, Appl  
Sequence 16, Appl





PT Polypeptides with heat-resistant cellobiohydrolase activity for efficient  
 PT breakdown of cellulose biomass.

XX  
 PS Disclosure; Page 45-47; 50pp; Japanese.

XX This invention describes a novel polypeptide originating in *Pyrococcus*  
 CC horikoshii OT3 which has cellobiohydrolase activity. The polypeptide of  
 CC the invention is capable of decomposing poly(D-glucopyranose) having beta  
 CC -1,4 bonds and can be used for the efficient and straightforward  
 CC breakdown of cellulose biomass to glucose. This sequence represents a P.  
 CC horikoshii OP3 cellobiohydrolase associated protein described in the  
 CC method of the invention

XX  
 SQ Sequence 458 AA;

Query Match 38.0%; Score 755; DB 3; Length 458;  
 Best Local Similarity 45.0%; Pred. No. 1,2e-62;  
 Matches 161; Conservative 46; Mismatches 111; Indels 40; Gaps 10;

QY 8 TSGREILDANNVVRAGINNFGETCNVYVHGLWSRDRYRSMLDQIKSLGNTIRLPYSD 67

DB 54 TSGER-----TPIHLFGVNWFGFETPNVHGLWKRNMEDMLQIKSLGFNAIRLPFCT 107

QY 68 DILKPGTWPNSINFYOMQDLOGLTSLOVMDKIYAYAGQIGLRILDRHRPDCSQSALW 127

DB 108 ESKVPTGTIGID-YSKNPDLRGDLSQIMEKIKKAGDLGIFVLVDYHRIIGCTHIEPLW 166

QY 128 YTSVSEATWISDQALQARVKGNTVVGFDLHNEPHD-----PACWCGDPS 175

DB 167 YTEDSEEDFINTWIEVAKRFGKYNNVIGADLKNEPHSVTSPPAAATDGTATWGMGNA 226

QY 176 IDWRLAARAGNAVLSVNPMLLI FVEGVQSYN-----GDSYTWGNNLQAGQYPVV 226

DB 227 TDWNLAAERIGAKILKVAHPHLIFVEGTQFTNPKTDSYKGYNAWGGNLMVAKYFV- 285

QY 227 LNVP-NRLVYSADHYATSVGPTWFS DPT-FPNNMGIWNNKNGYLFNQNIAPIVWLGEGF 284

DB 286 -NLPRKLVYSPHVYGPVYQYVFGPAKGFNDLPDIWYHFGYKLELGYSVWIGFEG 344

QY 285 TTL---QSTTDQTKTLVQYLRLPTAQYGADSFQWTFNSWNPDSGDTGGILKDDQVTVK 339

DB 345 GKYGHGDPDRVIWQNKLVDMW--IENKFCDF---YNSWNPDSGDTGGILQDDWTTI 397

RESULT 14

AAB96264

ID AAB96264 standard; protein; 514 AA.

XX  
 AC AAB96264;

XX  
 DT 29-OCT-2001 (first entry)

XX  
 DE Putative P. abyssi endoglucanase #2.

XX  
 KW Hyperthermophilic archaeon; hyperthermophilic protein.

XX  
 OS *Pyrococcus abyssi*.

XX  
 PN FR2792651-A1.

XX  
 PD 27-OCT-2000.

XX  
 PF 21-APR-1999; 99FR-00005034.

XX  
 PR 21-APR-1999; 99FR-00005034.

XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.

XX  
 PI (IFRE-) IFREMER INST FR RECH EXPL MER.

XX  
 PI Forterre P, Thierry JC, Priour D, Dietrich J, Lecompte O;

XX  
 PI Querellou J, Weissenbach J, Saurin W, Heilig R;

XX  
 PI WPI; 2001-126236/14.

XX  
 PT New nucleotide sequences isolated from *Pyrococcus abyssi* encode proteins  
 PT useful in industry.

XX  
 XX Claim 7; Page 921-922; 1657pp; French.

XX The present invention relates to the genomic sequence of *Pyrococcus*  
 CC abyssi (see AAF86431 and AAH41223-7) and P. abyssi proteins. P. abyssi is  
 CC a hyperthermophilic archaeon which is isolated from deep-sea  
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade. Note: This patent is in the same patent family as  
 CC WO200065062, which contains additional sequences as shown in AAB99132-  
 CC AAB99143, AAH75903-AAH75920 and AAG66436

XX  
 SQ Sequence 514 AA;

Query Match 34.3%; Score 682.5; DB 4; Length 514;

Best Local Similarity 41.6%; Pred. No. 1.1e-55;

Matches 146; Conservative 57; Mismatches 122; Indels 29; Gaps 10;

QY 7 HTSGREILDANNV-----PVRIAGINWFGFETCNVYVHGLWSRDRYRSMLDQIKSLGYNT 60

DB 30 YTAENGILFVQNVTTGEEKPLYLHGVSFWFGFELKDHVYGLDKRWKDKLDKVKRLGFNA 89

QY 61 IRLPYSDILKPGTNP--SINFYOMQDLOGLTSLOVMDKIYAYAGQIGLRILDRHRP 118

DB 90 IRLPFCSEISIRPDRSPERIN-YELNPDILKNTLSIEMEIKIIEYANSIGLYLLDYHRI 148

QY 119 DCSGOSALWYTSVSEATWISDQALQARVKGNTVVGFDLHNEPHDPAWCGDPSIDW 178

DB 149 GCEIEPLATYENYSEQYIKDWIFLAKRFGKYPNVIGADIKNEPHGEAGWGTGDER-DF 207

QY 179 RLAEERAGNAVLSVNPMLLI FVEGVQ-----SYNGDSYN--WGNLQAGQYPVVL 227

DB 208 RLFAEKVGRLEKVAHPHLIFVEGTQFTNPKTDSYKGYNAWGGNLMVAKYFVRL 267

QY 228 NVPRNLVYSAHDYATSVGPTWFS DPTFPNNMGIWNNKNGYLFNQNIAPIVWLGEGFTTL 287

DB 268 -PRGKVYSPHVYGPVYQYVFGPAKGFNDLPDIWYHFGYKLELGYSVWIGFEG 325

QY 288 QSTTDQTKTLVQYLRLPTAQYGADSFQWTFNSWNPDSGDTGGILKDDQVTVTVK 343

DB 326 BG-LDKVWQDAFVKWLLIKKIY-----NFFYWCINPESGDTGGIFLDDWKTVNWEK 375

RESULT 15

AAY13494

ID AAY13494 standard; protein; 616 AA.

XX  
 AC AAY13494;

XX  
 DT 30-JUL-1999 (first entry)

XX  
 DE Truncated cellulase Cel E3/B5.

XX  
 KW Cellulase; proteinase; truncated; Cel B5; Cel B4/5; Cel E1; Cel E1/2;

XX  
 KW Cel 1/2/3; Cel 6; Cel E3/B5; laundry detergent; stain-removing;  
 cotton-containing fabric; stonewashing.

XX  
 OS Unidentified.

XX  
 PN EP921188-A2.

XX  
 PD 09-JUN-1999.

XX  
 PF 15-SEP-1998; 96EP-00810919.

XX  
 PR 19-SEP-1997; 97US-00932571.

XX  
 PA (CLRN ) CLARIANT FINANCE BVI LTD.





KW emulsion stability; temperature resistance; pseudoplasticity; amylase;  
 KW cellulase; extracellular protease; intracellular protease;  
 KW glucose dehydrogenase; enzyme.  
 XX Xanthomonas campestris.  
 OS US2003036176-A1.  
 PN  
 XX  
 PD 20-FEB-2003.  
 XX  
 XX 10-AUG-2001; 2001US-00927827.  
 XX  
 XX 28-MAR-2001; 2001US-0279493P.  
 PR  
 XX (BOWE/) BOWER S G.  
 PA (RAMS/) RAMSEIER T M.  
 PA  
 XX  
 XX Bower SG, Ramseier TM;  
 PI  
 XX WPI; 2003-625389/59.  
 DR N-PSDB; ADD24896.  
 DR  
 XX New transformed cell or organism having reduced or enhanced activity of  
 PT at least one protein, useful for producing xanthan gum, which are useful  
 PT for providing formulations and properties.  
 PT  
 XX  
 PS Claim 1; SEQ ID NO 50; 135pp; English.  
 XX  
 CC The present invention relates to polypeptide and polynucleotide sequences  
 CC from Xanthomonas campestris which may be used for activity reduction or  
 CC enhancement using directed genetic engineering. A transformed cell or  
 CC organism having reduced or enhanced activity of at least one such protein  
 CC e.g. galactomannanase can be generated by disrupting the gene encoding  
 CC the protein. The activity of the protein is reduced by the presence of an  
 CC antisense nucleic acid sequence. The nucleic acid sequence of the gene  
 CC encoding the protein is a recombinant sequence having at least one  
 CC mutation as compared to the wild-type gene encoding the protein. The  
 CC transgenic cell or microorganism are useful for producing xanthan gum,  
 CC which are useful for providing formulations and properties, such as long-  
 CC term suspension and emulsion stability in alkaline, acid, and salt  
 CC solutions, temperature resistance, and pseudoplasticity. The present  
 CC sequence represents an enzyme relating to the present invention.  
 XX  
 SQ Sequence 535 AA;

Query Match 39.9%; Score 793.5; DB 7; Length 535;  
 Best Local Similarity 46.9%; Pred. No. 3.2e-66;  
 Matches 161; Conservative 58; Mismatches 105; Indels 19; Gaps 9;  
 QY 5 YNHTSGREILDANNVPRIAGINWFGFETCNVYVHGLMSRDRYSRMDQIKSLGYNTIRLP 64  
 DB 77 YSINNSRQIVDDSGKVQLKGVNVFGFETGNHVGELMARNNKDMVQMGGLGFNAVRLP 136  
 QY 65 YSDDILKPTMPNSINFYQMNQDQGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSQS 124  
 DB 137 FCFATLRSTMPASID-YSRNADQLGLTSLQLDKIVAEFNAGMYVLLDHTPDCAGIS 195  
 QY 125 ALWYTSVSEATWISDLOALAQRYKGNPTVVGFDLHNEPHDPACGCGDPSIDWRLAAER 184  
 DB 196 ELWYTSYTEAQWLADLRVANRYKNVPYVUGLDLKNPHGAAWTGTAATDNNKRAER 255  
 QY 185 AGNAVLNVNPNLLIFVEGVQ-----SYNGDSTWVGNGNLQAGQYVPVLNVP-NRLVYSAH 238  
 DB 256 GSAAVLAVAPKWLIAVEGHTDNPVCSTNG-GIFWGNLQPLACTP--LNI PANRLLALPH 312  
 QY 239 DYATSVGPOTWTSDFTPNNMPGINKWNGVLFNQIAPVWLGEFTTL--QSTTDQWL 296  
 DB 313 VYGPDPVYQSYFENDSNFNNMPAIWERHGFQAGTH--ALLGEFDGKYGEGDARDTKW 370  
 QY 297 KTLVQLRPTAQYADSFQWTFWSNPDSDGTGGILKDDWQTV 339  
 DB 371 DALVRYLR---SKGIN--QGFYWSNPNNSGDTGGILRDWTSV 408

RESULT 9  
 AAW34999  
 ID AAW34999 standard; protein; 628 AA.  
 XX  
 AC AAW34999;  
 XX  
 DT 27-AUG-2003 (revised)  
 DT 21-MAY-1998 (first entry)  
 XX  
 XX Archaeobacterium AEP11a endoglucanase.  
 XX  
 XX Endoglucanase; cellulase; carboxymethylcellulose; cellulose; biomass;  
 KW beta-1,4-glycosidic bond; hydrolysis; saccharification;  
 KW thermostable enzyme; thermophilic; glycosidase.  
 XX  
 OS archaeson.  
 OS Unidentified.  
 XX  
 FN W09744361-A1.  
 XX  
 PD 27-NOV-1997.  
 XX  
 PF 22-MAY-1997; 97WO-US008793.  
 XX  
 PR 22-MAY-1996; 96US-00651572.  
 XX  
 PA (RECO-) RECOMBINANT BIOCATALYSIS INC.  
 PA  
 PI Lam DE, Mathur EJ;  
 XX  
 DR WPI; 1998-018435/02.  
 DR N-PSDB; AAT94207.  
 XX  
 PT Endo:glucanase(s), preferably form archaeal bacterium, AEP11a - useful  
 PT to degrade carboxymethylcellulose and hydrolyse of beta-1,4-glycosidic  
 PT bonds in cellulose.  
 XX  
 PS Claim 1; Fig 10; 164pp; English.  
 XX  
 CC This protein comprises an endoglucanase of archaeobacterium AEP11a (Clone  
 CC 630P2), a hydrothermal vent isolate. The endoglucanase is capable of  
 CC degrading carboxymethylcellulose and of hydrolysing the beta-1,4-  
 CC glycosidic bonds in cellulose. It has homology to another endoglucanase  
 CC (see AAW34985) of archaeobacterium AEP11a. It can be produced from native  
 CC cells or from recombinant host cells, especially prokaryotic host cells  
 CC transformed with a plasmid or virus-derived vector including the  
 CC endoglucanase DNA (see AAT94207). 24 endoglucanases (see AAW34986-W35008)  
 CC are claimed. They can be used to degrade cellulose for the conversion of  
 CC plant biomass into fuels and chemicals, for use in detergents, textiles,  
 CC animal feed, waste treatment, and in the fruit juice and brewing  
 CC industries for the clarification and extraction of juices. (Updated on 27  
 CC -AUG-2003 to correct OS field.)  
 XX  
 SQ Sequence 628 AA;  
 Query Match 38.4%; Score 763; DB 2; Length 628;  
 Best Local Similarity 45.5%; Pred. No. 3.2e-63;  
 Matches 161; Conservative 48; Mismatches 109; Indels 36; Gaps 11;  
 QY 13 ILDANNVPIAGINWFGFETCNVYVHGLMSRDRYSRMDQIKSLGYNTIRLPYSSDILKP 72  
 DB 4 VAGTEETPIHLFGVNFPGFETPNYVHGLMSRWNEDMLLIQIKSLGFNARLPCTQSVKP 63  
 QY 73 GTPNSINFYQMNQDQGLTSLQVMDKIVAYAGQIGLRILDRHRPDCSQSALWYTSV 132  
 DB 64 GTWPTAID-YAKNPDLQGLDSVQIMEKIKAAGDLGIFVLVDYHRIGCNFIEPLWYDTSF 122  
 QY 133 SEATWISDLOALAQRYKGNPTVVGFDLHNEPHDP-----ACWCGDPSIDWRL 180  
 DB 123 SEQDYINTWYEVACRFKYNWVIGADLKNPEHSSSPAPAYTGDGSGATWGMGNATDNNL 182  
 QY 181 AAERAGNAVLNVNPNLLIFVEGVQ-----SYN-GDSYWWGNGNLQAGQYVPVLNVP- 230

CC activity comprises replacing an active site-associated amino glycosyl-  
CC stabilising amino acid with an amino acid that does not strongly bind a  
CC disaccharide product in the active site. Conversely, the method for  
CC making a glycosyl hydrolase with increased soluble substrate catalytic  
CC activity comprises replacing a hydrophobic substrate-binding amino acid  
CC with a positively charged residue. The invention also discloses mutants  
CC of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)  
CC produced according to the method of the invention. The Y245G mutant  
CC (AAB48788) has improved activity with insoluble substrates, and the W42R  
CC soluble substrates. The invention also encompasses DNA encoding these  
CC mutants. The glycosyl hydrolases of the invention are used as catalysts  
CC for cellulose hydrolysis to produce sugars that can be fermented to  
CC produce fuels such as ethanol. The present sequence represents the  
CC Acidothermus cellulolyticus E1 endoglucanase Y82R mutant  
XX  
SQ Sequence 521 AA;  
Query Match 99.1%; Score 1969; DB 4; Length 521;  
Best Local Similarity 99.4%; Pred. No. 7.2e-178;  
Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLDQIKSLGYNT 60  
DB 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLDQIKSLGYNT 60  
QY 61 IRLPYSDIILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDC 120  
DB 61 IRLPYSDIILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDC 120  
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
QY 181 AAB48786 standard; protein; 521 AA.  
DB 181 AAB48786;  
XX  
AC 09-MAR-2001 (first entry)  
XX  
DT  
XX  
DE Acidothermus cellulolyticus E1 endoglucanase mutant, W42R.  
XX  
XX E1 endoglucanase; glycosyl hydrolase; soluble substrate;  
XX cellulose hydrolysis; ethanol production; fermentation; mutant; muten.  
XX  
XX Acidothermus cellulolyticus.  
XX Synthetic.  
XX  
XX WO200070031-A1.  
XX  
XX 23-NOV-2000.  
XX  
XX 19-MAY-2000; 2000WO-US013971.  
XX  
XX 19-MAY-1999; 99US-0134925P.  
XX  
XX (MIDE ) MIDWEST RES INST.  
XX

PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;  
PI Decker SR;  
XX WPI; 2001-061226/07.  
XX Preparation of glycosyl hydrolase with an increased catalytic activity on  
XX insoluble substrate.  
PT  
XX Claim 16; Page 22-24; 30pp; English.  
XX  
XX The invention relates to a method for making glycosyl hydrolase mutants  
XX with increased catalytic activity with either insoluble or soluble  
XX cellulose substrates relative to the wild-type enzyme. The method for  
XX making a glycosyl hydrolase with increased insoluble substrate catalytic  
XX activity comprises replacing an active site-associated amino glycosyl-  
XX stabilising amino acid with an amino acid that does not strongly bind a  
XX disaccharide product in the active site. Conversely, the method for  
XX making a glycosyl hydrolase with increased soluble substrate catalytic  
XX activity comprises replacing a hydrophobic substrate-binding amino acid  
XX with a positively charged residue. The invention also discloses mutants  
XX of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788)  
XX produced according to the method of the invention. The Y245G mutant  
XX (AAB48788) has improved activity with insoluble substrates, and the W42R  
XX (AAB48786) and Y82R (AAB48787) mutants have improved activity with  
XX soluble substrates. The invention also encompasses DNA encoding these  
XX mutants. The glycosyl hydrolases of the invention are used as catalysts  
XX for cellulose hydrolysis to produce sugars that can be fermented to  
XX produce fuels such as ethanol. The present sequence represents the  
XX Acidothermus cellulolyticus E1 endoglucanase W42R mutant  
XX  
SQ Sequence 521 AA;  
Query Match 98.8%; Score 1964; DB 4; Length 521;  
Best Local Similarity 99.4%; Pred. No. 2.2e-177;  
Matches 356; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLDQIKSLGYNT 60  
DB 1 AGGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSLDQIKSLGYNT 60  
QY 61 IRLPYSDIILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDC 120  
DB 61 IRLPYSDIILKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGLRILDRHRPDC 120  
QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
QY 181 AAB48786 standard; protein; 521 AA.  
DB 181 AAB48786;  
XX  
AC 09-MAR-2001 (first entry)  
XX  
DT  
XX  
DE Xanthomonas campestris cellulase #3.  
XX  
XX Directed genetic engineering; galactomannanase; reduced activity;  
XX enhanced activity; xanthan gum production; suspension stability;  
XX

Db 282 ATSVVPTQWFSDFPTFPNNPGLIWKNGYLFNQNIAPVNLGEGFTLQSTTDDQWLKTLV 341

QY 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 358

Db 342 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 399

RESULT 5

AAI69508

ID AAY69508 standard; protein; 562 AA.

AC AAY69508;

XX 10-APR-2000 (first entry)

DT Acidotherrus cellulolyticus E1 endoglucanase.

XX E1 endoglucanase; cellulose binding domain; CBD; cellulose modification;

KW beta-1,4-endoglucanase; endocellulase; thermostable.

XX Acidotherrus cellulolyticus.

XX FH Key Location/Qualifiers

FT Peptide 1..41

FT Peptide /note= "Putative signal peptide"

FT Peptide 14..41

FT Domain /note= "Putative signal peptide (alternative)"

FT Domain 42..404

FT Region /note= "Catalytic domain"

FT Region 405..460

FT Domain /note= "Linker region"

FT Domain 461..562

FT /note= "Cellulose binding domain (CBD)"

XX CA2226898-AL.

XX PD 25-SEP-1999.

XX PF 25-MAR-1998; 98CA-02226898.

XX PR 25-MAR-1998; 98CA-02226898.

XX PA (MIDE ) MIDWEST RES INST.

XX PI Laymon RA, Adney WS, Thomas SR, Himmel ME;

XX WPI; 2000-087663/08.

XX N-PSDB; AA255924.

XX Isolated domains of Acidotherrus cellulolyticus E1 endoglucanase useful for labelling or modifying a cellulose and for purifying or immobilizing a binding domain fusion protein to cellulose.

XX Claim 1; Fig 2; 85pp; English.

XX This sequence represents the Acidotherrus cellulolyticus E1 endoglucanase, which is a beta-1,4-endoglucanase, or endocellulase. The cellulose binding domain (CBD) of E1 endoglucanase, and nucleotides which encode it are specifically claimed. The CBD is believed to be roughly wedge-shaped; as the CBD binds to cellulose, it is thought that the wedge tip is inserted between the microfibrils of the cellulose fibre, disrupting the crystalline structure, and making the cellulose linkages more accessible to the catalytic domain of the E1 endoglucanase. The E1 endoglucanase CBD is useful in labelling or modifying the surface of cellulose or other polysaccharides. Such modified cellulose can then be used in textile, pulp, paper, chemical and pharmaceutical industries. CBDs can be used in affinity purification of CBD-fusion proteins, and can also be used to immobilise the CBD-fusion proteins to a cellulose support. CBD-fusion proteins can be used to modify the chemical or physical properties of a cellulose or polysaccharide matrix column and to modify (e.g. roughen or disrupt) a cellulose or polysaccharide fibre. The CBD of the E1 endoglucanase exhibits greater stability at pH 4-8 and has an optimum temperature for stability of 83 degrees Celsius which is

CC not found in CBDs from non-thermophilic organisms

XX SQ Sequence 562 AA;

Query Match 99.3%; Score 1974; DB 3; Length 562;

Best Local Similarity 99.4%; Pred. No. 2.7e-178;

Matches 356; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGGYWHTSOREILDANNVPVRIAGINWFGFETCNVYVHGLNSRDYRSMIDQIKSLGYNT 60

Db 42 AGGGYWHTSOREILDANNVPVRIAGINWFGFETCNVYVHGLNSRDYRSMIDQIKSLGYNT 101

QY 61 IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 120

Db 102 IRLPYSDDILKPGTMPNSINFYQMNQDLQGLTSLQVMDKIVAYAGQIGLRIILDRHRPDC 161

QY 121 SQQSALWYTSVSEATWISDLQALAQRYKGNPVVGFDLHNEPHDPACWCGDPSIDWRL 180

Db 162 SQQSALWYTSVSEATWISDLQALAQRYKGNPVVGFDLHNEPHDPACWCGDPSIDWRL 221

QY 181 AAEERAGNAVLSVNPILLIFVEGVQSYNGDSYMWGGLQAGQYPPVNLVNPRLVYSAHDY 240

Db 222 AAEERAGNAVLSVNPILLIFVEGVQSYNGDSYMWGGLQAGQYPPVNLVNPRLVYSAHDY 281

QY 241 ATSVGPQWFSDFPTFPNNPGLIWKNGYLFNQNIAPVNLGEGFTLQSTTDDQWLKTLV 300

Db 282 ATSVGPQWFSDFPTFPNNPGLIWKNGYLFNQNIAPVNLGEGFTLQSTTDDQWLKTLV 341

QY 301 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 358

Db 342 QYLRPTAQYAGDSFQWTFWSWNPDSGDTGGILKDDQWQVDTVKDGYLAPIKSSIFDPV 399

RESULT 6

AAI69508

ID AAB48787 standard; protein; 521 AA.

XX AC AAB48787;

XX DT 09-MAR-2001 (first entry)

XX DE Acidotherrus cellulolyticus E1 endoglucanase mutant, Y82R.

XX KW E1 endoglucanase; glycosyl hydrolase; soluble substrate;

XX KW cellulose hydrolysis; ethanol production; fermentation; mutant; mutein.

XX OS Acidotherrus cellulolyticus.

XX OS Synthetic.

XX PN WO200070031-A1.

XX PD 23-NOV-2000.

XX PF 19-MAY-2000; 2000WO-US013971.

XX PR 19-MAY-1999; 99US-0134925P.

XX PA (MIDE ) MIDWEST RES INST.

XX PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;

XX PI Decker SR;

XX DR WPI; 2001-061226/07.

XX PT Preparation of glycosyl hydrolase with an increased catalytic activity on insoluble substrate.

XX PS Claim 17; Page 25-27; 30pp; English.

XX CC The invention relates to a method for making glycosyl hydrolase mutants with increased catalytic activity with either insoluble or soluble cellulose substrates relative to the wild-type enzyme. The method for making a glycosyl hydrolase with increased insoluble substrate catalytic

XX PR 15-JUL-1994; 94US-00276213.  
XX PA (MIDE ) MIDWEST RES INST.  
XX PI Thomas SR, Laymon RA, Himmel ME;  
XX DR WPI; 1996-105843/11.  
XX DR N-PSDB; AAT12337.  
XX PT New isolated DNA encoding endo:glucanase - obtd from Acidothermus  
XX PT cellulolyticus, used for prodn of the enzyme for use in cellulose  
XX PT hydrolysis.  
XX PS Claim 1; Page 22; 34pp; English.  
XX AC Acidothermus cellulolyticus E1 endoglucanase (AAR89927) is useful for  
CC hydrolyzing cellulosic biomass to sugars for simultaneous or subsequent  
CC fermentation to ethanol. It shows optimal activity at 83 deg C. The amino  
CC acid sequence of the mature enzyme was deduced from an E1 endoglucanase  
CC gene (AAT12337) cloned from A. cellulolyticus. Putative signal peptides  
CC were also identified (see also AAR89928-29). Cloning of this gene allows  
CC large-scale, low-cost prodn. of recombinant E1 endoglucanase, using pref.  
CC Saccharomyces, Zymomonas or E.coli hosts  
XX SQ Sequence 521 AA;  
Query Match 99.5%; Score 1978; DB 2; Length 521;  
Best Local Similarity 99.7%; Pred. No. 1e-178;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYSLMDQIKSLGYNT 60  
DB 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYSLMDQIKSLGYNT 60  
QY 61 IRLPYSDDLKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDC 120  
DB 61 IRLPYSDDLKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDC 120  
QY 121 SGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 121 SGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLQAGQYPPVVLNPNRLVYSAHDY 240  
DB 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLQAGQYPPVVLNPNRLVYSAHDY 240  
QY 241 ATSVGPOTWFSDDTFFNNMFGIWNKMGYLPNQNIAFVNLGEGTTLQSTTQDTWLTLY 300  
DB 241 ATSVGPOTWFSDDTFFNNMFGIWNKMGYLPNQNIAFVNLGEGTTLQSTTQDTWLTLY 300  
QY 301 QYLRPTAQYAGDSFQWTFWNNPDSGDTGGLKDDMQTVDTVKDGYLAPIKSSIFDPV 358  
DB 301 QYLRPTAQYAGDSFQWTFWNNPDSGDTGGLKDDMQTVDTVKDGYLAPIKSSIFDPV 358  
RESULT 4  
AAU79549  
ID AAU79549 standard; protein; 562 AA.  
AC AAU79549;  
XX 24-SEP-2002 (first entry)  
XX A. cellulolyticus cellulase E1 beta-1,4-endoglucanase precursor.  
XX DE Gene; ds; transgenic; plant; lignocellulose; cellulase; ligninase;  
KW fermentable sugar; ethanol; fermentation; silage; feed; fuel;  
KW industrial chemical; biodegradation; chloroaromatic;  
KW environmental pollutant; E1 beta-1,4-endoglucanase; ei.  
XX OS Acidothermus cellulolyticus.  
XX

FH Key Location/Qualifiers  
FT Peptide 1..41  
FT Protein /label= Leader\_sequence  
FT /label= Mature\_cellulase  
XX WO200234926-A2.  
XX 02-MAY-2002.  
XX 18-OCT-2001; 2001WO-US032538.  
XX 20-OCT-2000; 2000US-0242408P.  
XX (UNMS ) UNIV MICHIGAN STATE.  
XX Sticklen MB, Dale BE, Magbool S;  
XX WPI; 2002-489947/52.  
XX N-PSDB; ASK86729.  
XX Producing transgenic plants which after harvest degrade lignin and  
XX cellulose to fermentable sugars, by mating transgenic plant comprising  
XX DNA encoding cellulase with transgenic plant comprising DNA encoding  
XX ligninase.  
XX Disclosure; Page 96-99; 126pp; English.  
XX The invention discloses the production of a transgenic plant which  
XX degrades lignocellulose when the plant is grown. It comprises the  
XX production of the transgenic plant including cellulase and ligninase by  
XX mating a transgenic plant, containing a DNA encoding a cellulase, and a  
XX transgenic plant, containing a DNA encoding a ligninase, where both genes  
XX are operably linked to a nucleotide sequence encoding a signal peptide  
XX which targets the fusion protein to an organelle of the plant,  
XX particularly chloroplasts. The method is useful for producing a  
XX transgenic plant (e.g. maize) which degrades lignocellulose when the  
XX plant is grown to produce a plant material. This material is useful for  
XX converting lignocellulose, in a plant material, to fermentable sugars  
XX which are then fermented to ethanol. The transgenic plants also provide a  
XX plentiful and inexpensive source of fungal or bacterial cellulases and  
XX ligninases which can be used in the production of ethanol. They can also  
XX be used for pre-treating silage to increase the energy value of  
XX lignocellulosic feeds for cows and other ruminant animals, pre-treating  
XX lignocellulosic biomass for fermentation of chloroaromatic environmental  
XX pollutants. The protein sequence presented is the A. cellulolyticus  
XX cellulase E1 beta-1,4-endoglucanase precursor  
XX Sequence 562 AA;  
SQ  
Query Match 99.5%; Score 1978; DB 5; Length 562;  
Best Local Similarity 99.7%; Pred. No. 1e-178;  
Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYSLMDQIKSLGYNT 60  
DB 42 AGGYWHTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDRYSLMDQIKSLGYNT 101  
QY 61 IRLPYSDDLKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDC 120  
DB 102 IRLPYSDDLKPGTNPNSINFYQNMQLQGLTSLQVMDKIVAYAGQIGRLILDRHRPDC 161  
QY 121 SGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
DB 162 SGOSALWYTSVSEATWISDLQALAQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 221  
QY 181 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLQAGQYPPVVLNPNRLVYSAHDY 240  
DB 222 AAERAGNAVLNVNPNLLIFVEGVQSYNGDSYVWGGNLQAGQYPPVVLNPNRLVYSAHDY 281  
QY 241 ATSVGPOTWFSDDTFFNNMFGIWNKMGYLPNQNIAFVNLGEGTTLQSTTQDTWLTLY 300



CC (AAB48788) has improved activity with insoluble substrates, and the W42R  
 CC (AAB48786) and Y82R (AAB48787) mutants have improved activity with  
 CC soluble substrates. The invention also encompasses DNA encoding these  
 CC mutants. The glycosyl hydrolases of the invention are used as catalysts  
 CC for cellulose hydrolysis to produce sugars that can be fermented to  
 CC produce fuels such as ethanol. The present sequence represents the  
 CC Acidothermus cellulolyticus E1 endoglucanase Y245G mutant  
 XX  
 XX Sequence 521 AA;

Query Match 100.0%; Score 1987; DB 4; Length 521;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-179;  
 Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGGYWHSTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMMLDQIKSLGYNT 60  
 DB 1 AGGGYWHSTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMMLDQIKSLGYNT 60  
 QY 61 IRLPYSDDDILKPGTWPNSINFQMNQDLQGLTSLQVMDKIYVAYAGQIGLRILDRHRPDC 120  
 DB 61 IRLPYSDDDILKPGTWPNSINFQMNQDLQGLTSLQVMDKIYVAYAGQIGLRILDRHRPDC 120  
 QY 121 SQQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 121 SQQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 QY 181 AABRAGNAVLNPNLLIFVEGVQSYNGDSYMWGNLQAGQYPVVLNPNRLVYSAHDY 240  
 DB 181 AABRAGNAVLNPNLLIFVEGVQSYNGDSYMWGNLQAGQYPVVLNPNRLVYSAHDY 240  
 QY 241 ATSVGPQWFSDPTFPNNMPGIWNNKNGYLFNQNIAPVWLGEFGTTLOSTTDQTLWLTIV 300  
 DB 241 ATSVGPQWFSDPTFPNNMPGIWNNKNGYLFNQNIAPVWLGEFGTTLOSTTDQTLWLTIV 300  
 QY 301 QYLRTAQAGDSFQWTFWSNPNDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358  
 DB 301 QYLRTAQAGDSFQWTFWSNPNDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358

RESULT 2  
 AAW39262  
 ID AAW39262 standard; protein; 358 AA.

XX AAW39262;  
 XX  
 XX  
 XX 14-MAY-1998 (first entry)  
 XX  
 XX A. cellulolyticus E1-CAT translated region.  
 DE  
 DE E1 gene; endoglucanase; E1-CAT; catalytic domain; CD; enzyme; stability;  
 KW truncated; cellulose hydrolysis; biomass conversion.  
 KW

OS Acidothermus cellulolyticus.  
 XX  
 XX US5712142-A.  
 XX  
 XX PD 27-JAN-1998.  
 XX  
 XX PF 22-FEB-1996; 96US-00604913.  
 XX  
 XX PR 26-SEP-1989; 89US-00412434.  
 XX PR 27-JAN-1992; 92US-00826089.  
 XX PR 21-SEP-1993; 93US-00125115.  
 XX PR 15-JUL-1994; 94US-00276213.  
 XX  
 XX PA (MIDE ) MIDWEST RES INST.  
 XX  
 XX PI Chou Y, Himmel ME, Baker JO, Thomas SR, Adney WS;  
 XX  
 XX DR WPI; 1998-119985/11.  
 XX DR N-PSDB; AAV09659.  
 XX

PT DNA encoding truncated form of E1 endoglucanase from Acidothermus

PT cellulolyticus - and related vectors and transformed cells, expressing  
 PT only catalytic domain, is used for biomass conversion and has better heat  
 PT stability than complete enzyme.

XX  
 XX Claim 2; Fig 5; 19pp; English.  
 XX  
 XX This sequence represents a novel Acidothermus cellulolyticus endoglucanase  
 CC E1 protein (E1-CAT) which expresses only the catalytic domain (CD) of the  
 CC enzyme, without its peptide linker or cellulose binding domain (CBD) and  
 CC the truncated enzyme expressed by is used for hydrolysis of cellulose  
 CC (biomass conversion). Compared with full-length E1, the truncated enzyme  
 CC has better heat stability and higher temperature of maximum activity  
 XX  
 XX Sequence 358 AA;

Query Match 99.5%; Score 1978; DB 2; Length 358;  
 Best Local Similarity 99.7%; Pred. No. 5.8e-179;  
 Matches 357; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGGGYWHSTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMMLDQIKSLGYNT 60  
 DB 1 AGGGYWHSTSGREILDANNVPVRIAGINWFGFETCNVYVHGLWSRDYRSMMLDQIKSLGYNT 60  
 QY 61 IRLPYSDDDILKPGTWPNSINFQMNQDLQGLTSLQVMDKIYVAYAGQIGLRILDRHRPDC 120  
 DB 61 IRLPYSDDDILKPGTWPNSINFQMNQDLQGLTSLQVMDKIYVAYAGQIGLRILDRHRPDC 120  
 QY 121 SQQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 DB 121 SQQSALWYTSVSEATWISDLQALQRYKGNPTVVGFDLHNEPHDPACWCGDPSIDWRL 180  
 QY 181 AABRAGNAVLNPNLLIFVEGVQSYNGDSYMWGNLQAGQYPVVLNPNRLVYSAHDY 240  
 DB 181 AABRAGNAVLNPNLLIFVEGVQSYNGDSYMWGNLQAGQYPVVLNPNRLVYSAHDY 240  
 QY 241 ATSVGPQWFSDPTFPNNMPGIWNNKNGYLFNQNIAPVWLGEFGTTLOSTTDQTLWLTIV 300  
 DB 241 ATSVGPQWFSDPTFPNNMPGIWNNKNGYLFNQNIAPVWLGEFGTTLOSTTDQTLWLTIV 300  
 QY 301 QYLRTAQAGDSFQWTFWSNPNDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358  
 DB 301 QYLRTAQAGDSFQWTFWSNPNDSGDTGGILKDDQVTDVTKDGYLAPIKSSIFDPV 358

RESULT 3  
 AAR89927  
 ID AAR89927 standard; protein; 521 AA.

XX AAR89927;  
 XX  
 XX 08-OCT-1996 (first entry)  
 XX  
 XX A. cellulolyticus E1 endoglucanase.  
 DE  
 DE E1 endoglucanase; cellulase; cellulose; saccharification; ethanol.  
 KW  
 KW Acidothermus cellulolyticus.  
 OS  
 OS Key Location/Qualifiers  
 FH Domain 1..363  
 FT /label= Catalytic-domain  
 FT Domain 364..417  
 FT /label= Linker  
 FT /note= "proline/serine/threonine-rich linker domain  
 FT common to multi-domain microbial cellulases"  
 FT Domain 418..521  
 FT /label= Cellulose-binding\_domain

XX WO9602551-A1.  
 XX  
 XX 01-FEB-1996.  
 XX  
 XX 14-JUL-1995; 95WO-US008868.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 2, 2004, 16:29:04 ; Search time 47.6667 Seconds  
(without alignments)  
2122.070 Million cell updates/sec

Title: US-09-997-504A-10

Perfect score: 1987  
Sequence: 1 AGGVTHTSGREILDANNVP.....VDTKDGVLAPIKSIIDPV 358

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1987	100.0	521	4	AAB48788
2	1978	99.5	358	2	AAB39262
3	1978	99.5	521	2	AAR89927
4	1978	99.5	562	5	AAB79549
5	1974	99.3	562	3	AAB69508
6	1969	99.1	521	4	AAB48787
7	1964	98.8	521	4	AAB48786
8	793.5	39.9	535	7	ADD24922
9	763	38.4	628	2	AAB34999
10	763	38.4	841	2	AAB34985
11	762	38.3	553	5	ABG70759
12	755	38.0	430	3	AAB10344
13	755	38.0	458	3	AAB10345
14	682.5	34.3	514	4	AAB36284
15	621	31.3	616	5	AAB13494
16	621	31.3	616	5	AAB16325
17	621	31.3	1426	2	AAB13492
18	621	31.3	1426	5	AAB16323
19	513.5	25.8	425	4	ABD09060
20	276.5	13.9	582	7	ADD24920
21	267	13.4	508	5	ABD24433
22	211.5	10.6	488	5	ABD24433
23	205	10.3	38	2	AAB99930
24	195	9.8	389	5	AAB49427
25	185	9.3	517	6	ABP99336

26	183	9.2	551	5	ABB92442
27	173.5	8.7	782	6	ABP73022
28	171.5	8.6	375	6	ABP73019
29	160.5	8.1	722	6	ABU24069
30	156.5	7.9	518	4	ABB06928
31	153	7.7	420	6	ABP96833
32	147.5	7.4	406	2	AAR08199
33	146	7.3	335	7	ADD24924
34	143	7.2	400	4	AAE99785
35	141.5	7.1	329	2	AAW35002
36	141	7.1	490	7	ADC58031
37	140	7.0	431	3	AAG42172
38	140	7.0	431	3	AAG27501
39	140	7.0	431	5	ABB93972
40	140	7.0	442	3	AAG42171
41	140	7.0	443	3	AAG27500
42	139	7.0	348	4	AAE99788
43	139	7.0	476	3	AAAY54123
44	139	7.0	490	3	AAAY54122
45	138	6.9	335	4	AAE12786

#### ALIGNMENTS

RESULT 1  
AAB48788  
ID AAB48788 standard; protein; 521 AA.

XX AC AAB48788;

XX DT 09-MAR-2001 (first entry)

XX DE Acidothermus cellulolyticus E1 endoglucanase mutant, Y245G.

XX KW E1 endoglucanase; glycosyl hydrolase; insoluble substrate;

XX KW cellulose hydrolysis; ethanol production; fermentation; mutant; mutein.

XX OS Acidothermus cellulolyticus.

XX OS Synthetic.

XX PN WO200070031-A1.

XX PD 23-NOV-2000.

XX PF 19-MAY-2000; 2000WO-US013971.

XX PR 19-MAY-1999; 99US-0134925P.

XX XX (MIDE ) MIDWEST RES INST.

XX PI Himmel ME, Adney WS, Baker JO, Vinzant TB, Thomas SR, Sakon J;

XX PI Decker SR;

XX DR WPI; 2001-061226/07.

XX PT Preparation of glycosyl hydrolase with an increased catalytic activity on insoluble substrate.

XX CC Claim 5; Page 27-29; 30pp; English.

XX CC The invention relates to a method for making glycosyl hydrolase mutants with increased catalytic activity with either insoluble or soluble cellulose substrates relative to the wild-type enzyme. The method for making a glycosyl hydrolase with increased insoluble substrate catalytic activity comprises replacing an active site-associated amino glycosyl-stabilising amino acid with an amino acid that does not strongly bind a disaccharide product in the active site. Conversely, the method for making a glycosyl hydrolase with increased soluble substrate catalytic activity comprises replacing a hydrophobic substrate-binding amino acid with a positively charged residue. The invention also discloses mutants of Acidothermus cellulolyticus E1 endoglucanase (AAB48786-B48788) produced according to the method of the invention. The Y245G mutant